8-22-2019

Operational Decision Making under Uncertainty: Inferential, Sequential, and Adversarial Approaches

Andrew J. Keith

OPERATIONAL DECISION MAKING UNDER UNCERTAINTY: INFERENTIAL, SEQUENTIAL, AND ADVERSARIAL APPROACHES

DISSERTATION

Andrew J. Keith, Capt, USAF

AFIT-ENS-DS-19-S-041

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-DS-19-S-041

OPERATIONAL DECISION MAKING UNDER UNCERTAINTY: INFERENTIAL, SEQUENTIAL, AND ADVERSARIAL APPROACHES

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in Operations Research

Andrew J. Keith, MS

Capt, USAF

September 2019

AFIT-ENS-DS-19-S-041

OPERATIONAL DECISION MAKING UNDER UNCERTAINTY: INFERENTIAL, SEQUENTIAL, AND ADVERSARIAL APPROACHES

DISSERTATION

Andrew J. Keith, MS
Capt, USAF

Committee Membership:

Darryl K. Ahner, PhD
Chair

Raymond R. Hill, PhD
Member

Brian J. Lunday, PhD
Member

Lt Col Richard S. Seymour, PhD
Member

Adedeji B. Badiru, PhD
Dean, Graduate School of Engineering and Management

Abstract

Modern security threats are characterized by a stochastic, dynamic, partially observable, and ambiguous operational environment. This research addresses decision making under uncertainty in operations planning, analysis, and assessment for such complex security threats. First, a review of the literature on uncertainty modeling, decision making, and optimization under uncertainty focuses on recent advances in ambiguity modeling and optimization practice. This review provides a framework for the subsequent methodological and applied research and provides a comprehensive review of contemporary applications of decision making and optimization under uncertainty in the literature. Next, a survey of uncertainty models for military assessment addresses both qualitative and mixed-method approaches to complement the quantitative models discussed in the literature review. This survey provides a research-based guide for practitioners to apply qualitative but rigorous uncertainty models to practical assessment problems.

Following the reviews of existing literature and practice, this research develops a new method for decision making under uncertainty in an inference setting. A method for robust queue inference addresses a general class of queues in which the internal queueing system is unobservable and the departure and arrival times are stochastic and partially observable. This work improves decision makers' ability to analyze queues in uncertain environments using a principled method that provably converges to the true parameter value and has strong empirical performance.

Next, the research transitions from inference to sequential decision making with an original formulation and solution method for robust information collection in dynamic, partially observable, and ambiguous environments. The solution method has desirable

theoretical convexity and convergence properties. A computational experiment shows improved performance compared to existing methods for a set of classic problems from the literature. In addition, a detailed application to a cybersecurity detection problem illustrates the efficacy of the new formulation and solution method.

Lastly, a new application of optimal and approximate techniques for solving large-scale, extensive-form games with imperfect information addresses the dynamic, stochastic, and partially observable multi-agent environment. This work provides explicit details for optimal and approximate formulations of a multi-domain cyber and air defense problem, produces near-optimal strategies, characterizes the optimality gap of the approximate solutions, and analyzes the sensitivity of results to key problem parameters. Additionally, an extension to robust opponent exploitation incorporates bounded rationality and model ambiguity. The robust formulation addresses both the cyber-physical nature of the problem and the adversarial uncertainty. Empirical evidence demonstrates the effectiveness of the robust approach when the opponent plays with bounded rationality.

Collectively, these contemporary surveys, methodological advances, and new applications provide a suite of mathematical tools and computational algorithms for solving complex decision making problems under uncertainty in challenging settings. This research improves the capabilities of decision making and optimization by capturing the state of the art and practice and by extending existing algorithms to ambiguous and partially observable settings.

To my family.

Acknowledgements

This work would not have been possible without the support of many mentors, colleagues, and family members. First, I would like to thank my advisor Dr. Darryl Ahner for the academic and professional guidance over the past three years. Your encouragement to keep pushing the boundaries of our research led to rewarding work in an exciting but challenging field. I appreciate the many hours you dedicated to discussing our work and reviewing drafts despite a full schedule of competing priorities across the institute. I also appreciate your professional guidance and flexibility in helping me to accomplish my military duties in addition to these academic efforts.

Next, I would like to thank my committee members, Dr. Raymond Hill, Dr. Brian Lunday, and Lt Col Richard Seymour, for their support and feedback. Your academic instruction gave me a solid foundation for the work I wanted to pursue and your feedback was instrumental in improving my research. I also appreciate your encouragement throughout the program when the light at the end of the tunnel sometimes seemed far away.

I would also like to thank the broader Military Operations Research Society community. I received valuable feedback at the MORS symposia and I was exposed to many interesting ideas. In particular, I would like to thank LTC Marv King for his boundless energy dedicated to professionalizing the operation assessment community and for supporting my assessment research. I would also like to thank LTC Nicole Curtis for her critical operational perspective during our collaboration on operation assessment research. The MORS community provides an invaluable home for analysts across the Department of Defense and related national security organizations.

Although I was the last to join, special thanks go to my cohort for their professional contributions to this work. Without you, these past few years would have been much less enjoyable. It has been great to see our families grow together and to share the

ups and downs of research in the office. MAJ Alexander Kline, Capt Nick Caballero, and Capt Phil Jenkins: thank you for the companionship and willingness to help out with professional questions no matter how pressing your other work was. I feel lucky to have had the opportunity to work with each of you.

Thank you all for your support and professional advice.

Andrew J. Keith

Table of Contents

Page

Page

Page

List of Figures

List of Tables

OPERATIONAL DECISION MAKING UNDER UNCERTAINTY:
INFERENTIAL, SEQUENTIAL, AND ADVERSARIAL APPROACHES

I. Introduction

The United States Air Force Future Operating Concept (2015a) envisions a challenging security environment in 2035. Adversaries may have the capability to conduct fully integrated, multi-domain operations in a way that achieves disproportionately destructive effects. Consider the increased difficulty of deterring and responding to not only a ballistic missile attack, but also to an integrated information campaign and cyber infiltration. The synergistic effects of this evolution in combined arms warfare necessitates improvements in both operational art and operational science.

The concerns introduced in the multi-domain attack vignette extend to the broader security community, as emphasized in the 2018 National Defense Strategy (Mattis, 2018). The security environment is undergoing rapid technological innovation and is becoming increasingly complex and uncertain. Complex environments and adaptive adversaries create fundamental limits on the ability to predict future outcomes, especially at strategic levels. However, in many planning and assessment problems, the operating environment is complex, but not so complex that there is a complete lack of information for decision making. The environment is also rarely so clearly defined and understood that the uncertainty is negligible. Decision makers face a dynamic environment characterized by thinking adversaries and varying levels of uncertainty, ambiguity, and partial observability. How should decision makers address such a challenging environment?

While rapid technological change introduces security challenges, it also presents

new opportunities. Increasing computing power has complemented foundational improvements in statistical, simulation, and optimization algorithms to improve the capabilities of quantitative methods for decision making problems. Operations research methods and closely related techniques from applied mathematics, statistics, computer science, machine learning, and artificial intelligence can now address complex decision problems under uncertainty on a practical scale. While some of the most important and difficult security problems remain outside the scope of quantitative algorithms, there is a strategic imperative to exploit emerging solution techniques to expand the class of problems for which modern operations research methods are tractable.

Recent advances in these methods have focused on exploiting partial knowledge in a variety of complex settings. It is rarely the case that the operational environment is either completely known or completely unknown. Instead, the decision maker encounters uncertainty due to partially observable states and rewards, ambiguous transition dynamics, and intelligent adversaries. Exploiting partial information about the operational environment enables decision makers to improve decisions by embracing uncertainty and developing solutions that are robust to rapidly evolving environments and adversaries. This dissertation focuses on solving decision problems characterized by stochastic, partially observable, and ambiguous environments in static, dynamic, and multi-agent settings.

The dissertation is structured as a series of independent scholarly articles addressing the topic of decision making under uncertainty in operations planning, analysis, and assessment. Chapter II reviews the literature on decision making and optimization under uncertainty, with a focus on recent advances in ambiguity models and optimization practice. Chapter III complements the theoretical review of quantitative literature in Chapter II with a survey of qualitative uncertainty and applications

to military assessment practice. Building on the foundational literature in Chapters II and III, Chapters IV, V, and VI explore decision making under uncertainty in static, dynamic, and multi-agent settings, respectively. Chapter IV develops a new method for robust queue inference with partially observable, stochastic arrival and departure times. This general method is applicable to arbitrary queues but is motivated specifically by cybersecurity and terrorism applications. Chapter V develops a new method for robust information collection in dynamic, partially observable and ambiguous environments, with an extended application to a cybersecurity detection problem. Chapter VI presents a new application to a multi-domain cyber and air defense problem using optimal and approximate techniques for solving extensive-form games with imperfection information.

1.1 Summary

In particular, Chapter II provides a review of the theoretical foundations for the methodological and applied research in the subsequent chapters. Recent advances in decision making have incorporated both risk and ambiguity in decision making models and optimization methods. These methods implement a variety of uncertainty representations from probabilistic and non-probabilistic foundations, including traditional probability theory, uncertainty sets, ambiguity sets, possibility theory, evidence theory, fuzzy measures, and imprecise probability. The choice of uncertainty representation impacts the expressiveness and tractability of the decision models. This chapter surveys recent approaches for representing uncertainty in both decision making and optimization to clarify the trade-offs between the alternative representations. Robust and distributionally robust optimization are surveyed, with particular attention to standard form ambiguity sets. Applications of uncertainty and decision models are also reviewed, with a focus on recent optimization applications.

Chapter III complements the quantitative work in Chapter II with a focused survey of uncertainty models in military assessment. Evaluation theory provides a rigorous foundation for the practice of military operation assessment. Government and industry assessors have used evaluation theory to improve the effectiveness of assessment across a wide range of fields. This chapter focuses on the relationship between evaluation theory and military assessment. The chapter briefly surveys the major evaluation approaches with a focus on connecting the theoretical models to practical, security-related applications. These evaluation approaches include expertise-oriented, program-oriented, decision-oriented, and participant-oriented models. Within the overarching framework of these approaches, alternative monitoring and evaluation designs are considered in detail, including descriptive designs (e.g., case study, cross-sectional, time-series), quasi-experimental designs (e.g., interrupted time-series, comparison group, case study), and experimental designs (e.g., posttest-only, pre-post). Then, the chapter discusses quantitative and qualitative methods for analyzing and reporting uncertainty with respect to each design alternative, with an emphasis on mixed-method approaches. Throughout the chapter, applied examples make explicit the relationship between evaluation theory and operation assessment practice.

Chapter IV develops a new method for robust queue inference. The internal structure and parameters of a queue are completely unobservable in some military and competitive commercial applications. Furthermore, arrival and departure times may be observable but subject to substantial uncertainty due to measurement error in an adversarial environment. This analysis estimates the number of servers in an internally unobservable, first-come, first-served G/G/c queue using an order-based approach. This new approach provides a lower bound and converges in probability to the correct value. Compared to the standard variance minimization method, the order-based approach has improved performance for small samples. The order-based

algorithm is robust to noise in arrival and departure time measurements, whereas the variance minimization approach exhibits poor performance with noisy data. An extension to last-come, first-served G/G/c queues is also considered. The last-come, first-served order-based approach also provides a lower bound that converges in probability to the correct number of servers.

Chapter V develops a new formulation and method for robust solutions to partially observable Markov decision processes (POMDPs) with ambiguous transitions and belief-rewards. The chapter introduces robust belief-reward partially observable Markov decision processes as a generalization of Markov decision processes that allows for state uncertainty, model uncertainty, and belief-dependent rewards. In many practical applications, POMDP transition and observation parameters are difficult to estimate. This research shows that traditional POMDP solution techniques are highly sensitive to model misspecification, particularly in the belief-reward setting. To address this challenge, a robust belief-reward algorithm is developed that extends point-based value iteration while retaining desirable flexibility and convergence properties. In addition to the foundational theoretical properties, an empirical investigation shows that the robust solution technique provides protection against model misspecification in several different problem classes. To illustrate the importance of addressing model misspecification for information acquisition problems, this chapter also presents an application of the robust belief-reward POMDP formulation to a cybersecurity problem which demonstrates improved performance under worst-case dynamics.

Chapter VI presents a new application of optimal and approximate extensive-form solution techniques to an integrated cyber and air defense problem with imperfect information. Emerging multi-domain threats require an integrated defensive strategy. This chapter develops multi-domain security games to address the combined

cyber-physical threat to national population centers. This research uses a zero-sum, extensive-form game to model an attacker and defender in physical and cyber space, drawing on both the cybersecurity and ballistic missile defense literature to inform the game structure. To determine optimal defender strategies, a multi-domain security game is developed with a reformulation of the problem to find a Nash equilibrium using an efficient, sequence-form linear program. This chapter also develops an application of the approximate counterfactual regret minimization algorithm to this problem and characterizes the optimality gap. Furthermore, this research quantifies the value of improved situational awareness in the cyber domain and presents an extension to robust opponent exploitation.

1.2 Contributions

The literature reviews, methodological advances, and applications developed in this body of work contribute to the general field of operations research and specifically to military and security operations research practice. The literature review of decision making and optimization under uncertainty (Chapter II) organizes the disparate literature on theoretical uncertainty models, decision models, and optimization models into a coherent structure and identifies the relationship between the three research areas. In addition, the survey of uncertainty models in military assessment (Chapter III) contributes to military operations research practice by providing a research-based guide for practitioners to apply qualitative but rigorous uncertainty models to assessment problems.

Building on the foundation of this existing literature, a new method for robust queue inference contributes to the field by improving decision makers' ability to analyze queues in uncertain environments (Chapter IV). This server estimation method is valid for a large class of general queues with limited knowledge of the structure of the

queue and small, noisy samples for arrivals and departures. This research proves that the method produces an estimate which has theoretical convergence and lower-bound guarantees. It also presents empirical evidence for improved performance compared to existing methods across a broad range of parameter settings.

Extending to sequential decision making, an original robust belief-reward POMDP formulation and a newly developed solution algorithm contribute to the field by providing tools for solving a new class of information-collection problems under model ambiguity (Chapter V). This research proves that the solution technique has theoretical convexity and convergence properties that make it compatible with a mature family of approximation techniques. It also presents empirical evidence for improved performance compared to existing methods for a set of classic problems from the literature and for a practical cybersecurity detection problem in an ambiguous environment.

Lastly, an application of counterfactual regret minimization to a multi-domain cyber and air defense problem contributes to the literature by solving a contemporary operational problem with fast, near-optimal techniques (Chapter VI). This new application provides explicit details of optimal and approximate formulations for the problem and characterizes the optimality gap and sensitivity to key problem parameters in the multi-domain security setting, which differ significantly from other recent applications. It also presents an original robust formulation that addresses both the cyber-physical nature of the problem and the adversarial uncertainty. Empirical evidence demonstrates the effectiveness of the robust approach when the opponent plays with bounded rationality.

In addition to contributing to the literature through the publication of surveys, methods, applications, and results, this dissertation contributes open source software implementations for all methods and publishes raw data for all results. These code

and data products are available with testing, benchmarking, and documentation at `https://github.com/ajkeith`.

II. Literature Review

2.1 Introduction

Kolmogorov's axiomatic development of a rigorous, mathematical theory of probability (1933) is the dominant model for uncertainty in academic theory and practical application. Over the past several decades, various theories have been developed that generalize the theory of probability to address aspects of uncertainty that are difficult or impossible to model in standard probability theory. One of the major motivations for these models is to develop a practical and mathematically sound foundation for representing imprecision and ambiguity. In addition to the theoretical development of uncertainty models, the decision-theoretic literature has extended classical theories to uncertain and ambiguous settings. Furthermore, the recent attention to stochastic and robust models in the operations research literature has led to rapid growth in methods to model uncertainty in large-scale and practical decision problems. These models vary in terms of tractability and expressiveness.

This chapter reviews the most notable models for representing uncertainty, making decisions under uncertainty, and optimizing under uncertainty. We describe a framework that captures the relationships among the different uncertainty representations and decision models. We also discuss the connections between the representation, decision, and optimization models to clarify the structure of existing modeling approaches and to highlight promising areas for future research. To highlight the application of these models to current, practical problem solving, we emphasize recent advances for addressing uncertainty in the optimization literature.

In the following material, we introduce and compare the major uncertainty models in Section 2.2. Then, we discuss the main classes of decision making under uncertainty in Section 2.3 and the main classes of optimization under uncertainty in Section

2.4. Section 2.5 provides examples of applications, with a focus on recent optimization practice. Section 2.6 concludes by summarizing trends in decision making and optimization under uncertainty and identifying research opportunities.

2.2 Uncertainty Models

Uncertainty models have developed across distinct fields, resulting in a fragmented terminology. Although we present each model using its specific terminology, we generally follow the terminology used by Camerer and Weber (1992) to define three *degrees of uncertainty*. First, *certainty* refers to complete knowledge, or probability one, that an event will occur. Next, *risk* refers to probabilistic uncertainty with a known distribution, as discussed by Knight (1921). Risk is also referred to as aleatoric uncertainty (Ferson et al., 2004), first-order uncertainty (Klibanoff et al., 2005), or state uncertainty (Marinacci, 2015). Lastly, *ambiguity* refers to uncertainty with an unknown distribution, as discussed by (Ellsberg, 1961). Ambiguity is also referred to as epistemic uncertainty (Ferson et al., 2004), second-order uncertainty (Klibanoff et al., 2005), and model uncertainty (Marinacci, 2015). Ambiguity is also sometimes used interchangeably with uncertainty, without any qualification (Etner et al., 2012).

There are several different approaches to modeling uncertainty. In the degenerate case, deterministic models ignore uncertainty. Under the standard approach, probability distributions are used to model uncertainty stochastically. However, the decision maker may not have complete information about the probability distribution. This motivates second-order models of uncertainty, including sets of probability distributions. Although third and higher-order uncertainty models are possible, they are not common in the literature.

In addition to these fundamentally probabilistic models, there are a variety of models that can be developed without a probability theory foundation. This cate-

gory includes possibility theory, evidence theory, fuzzy measure theory, and imprecise probability. We consider each of these models in detail in the following subsections.

2.2.1 Probability Theory

Probability distributions are the standard representation of uncertainty. Standard probability theory is developed axiomatically by Kolmogorov (1933). Probability theory can represent most aspects of uncertainty that affect decision making and is built from a measure theory foundation. For all definitions in this section, we adjust notation to facilitate comparison but also present the terminology unique to each model, when appropriate.

Definition 1. (Kolmogorov, 2018) Given a non-empty set Ω and a σ-algebra $\mathcal{F}$, a *probability measure* is a function $P : \mathcal{F} \to [0, 1]$ such that the following axioms hold,

1. For all $E \in \mathcal{F}, P(E) \geq 0$
2. $P(\Omega) = 1$
3. Any countable sequence of mutually exclusive events $E_1, E_2, \ldots$ satisfies $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

The axioms in Definition 1 have been relaxed in various alternative, but similar, versions of probability theory (Fishburn, 1986). In the subjectivist tradition, de Finetti (1974) considers $A \cup B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$ in place of Definition 1, Axiom 3. Good (1977) noted that de Finetti regards Axiom 3 "as irrelevant for practical purposes and unjustifiable on theoretical and conceptual grounds." This alternative axiomatization replaces *countable additivity* with *finite additivity*. Subjectivity and countability can be assumed independently (Fishburn, 1986).

Probability theory is a special case of most other theories of uncertainty, which is not surprising since the intent of those theories is often to generalize standard proba-

bility. In some cases, this relationship is explicit (e.g., upper and lower probabilities). In other cases, the structure of both theories allow for such an interpretation, but it is not necessary for the development of either theory (e.g., transferable belief models). Although finitely additive probability has received some attention in the literature, it is used less frequently than countably additive probability, likely due to the added complexity.

2.2.2 Sets of Probability Measures

Sets of probability measures model uncertainty with an arbitrary set of probability measures, which can be finite or infinite. This category of uncertainty models includes *uncertainty sets* and *ambiguity sets* as major sub-classes. Uncertainty sets are used to model scenarios in which the only information about the value of a variable is whether or not it belongs to a given set. This approach implies a deterministic perspective on uncertainty in the sense that there is no probabilistic information available about the unknown variable.

Let $\mathcal{P}$ be the set of all probability measures on a given sample space Ω and σ-algebra $\mathcal{F}$, $\mathbb{P}$ be a probability measure, and $F_{\mathbb{P}}$ be the cumulative distribution function of $\mathbb{P}$. An ambiguity set, $\mathcal{M}$, can represent an uncertainty set, $\mathcal{C}$, for an unknown vector, $\tilde{\boldsymbol{z}}$, by including all compatible probability distributions: $\mathcal{M} = \{\mathbb{P} : \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{C}] = 1\}$ (Wiesemann et al., 2014). Conversely, in the degenerate case, an ambiguity set can represent a single distribution, $\mathcal{M} = \{\mathbb{P}\}$, which is interpreted as a stochastic or risk model where the distribution is known. In most applications, however, ambiguity sets are used to represent an intermediate level of knowledge between a single known distribution and a complete lack of probabilistic knowledge in the uncertainty set formulation.

Ambiguity sets generalize uncertainty sets by considering sets of probability distri-

butions rather than sets of real-valued vectors. Ambiguity sets can be interpreted as uncertainty sets on probability distributions or second-order uncertainty (e.g., Snow (2010)). Ambiguity sets are often closed and convex sets, in which case they are also referred to as *credal sets* and have a one-to-one relationship with coherent lower previsions (Walley, 2000; Augustin et al., 2014).

2.2.3 Possibility Theory

Fuzzy set theory introduces the notion of graded set membership, represented by a continuous number. This contrasts with traditional set theory, in which each element either is or is not a member of a given set. Formally, Zadeh (1965) defines fuzzy sets in the following way.

Definition 2. (Zadeh, 1965) Let the universe of discourse, Ω, be a set with a generic element of Ω denoted by ω so that $\Omega = \{\omega\}$. A *fuzzy set* X is characterized by its universe of discourse, Ω, and a membership function $f_X(\omega) : \Omega \to [0, 1]$ which associates each point in Ω with a real number in the interval $[0, 1]$. The value of $f_X(\omega)$ at ω represents the 'grade of membership' of ω in X.

Fuzzy logic is the specific application of fuzzy sets to logic statements. In this case, the membership function is interpreted as a truth value, with 0 representing completely false, 1 representing completely true, and intermediate values representing statements that are not completely true or false. This approach leads to multivalued logic with continuous truth values (Zadeh, 1965).

Traditional set theory is a special case of fuzzy set theory where $f_X(\omega) = 1$ or 0 (Zadeh, 1965). Similarly, traditional logic is a special case of fuzzy logic where truth values are binary. The membership function is somewhat similar to a scaled probability function, but the similarity does not hold when considering membership to

different sets (Zadeh, 1965). This prevents the membership function from being interpreted as a probability or belief that an element belongs to a set.

Fuzzy sets and fuzzy logic are the foundation for more complex "fuzzy" theories which are sometimes generically referred to as fuzzy logic, e.g., in fuzzy control applications (Maiers and Sherif, 1985). Zadeh et al. (2005) also propose a generalized theory of uncertainty that explicitly supports possibility theory, probability theory, fuzzy probability theory, fuzzy logic, fuzzy set theory, fuzzy graph theory, random sets, and the Dempster-Shafer theory of evidence (Zadeh et al., 2005). The generalized theory of uncertainty is a framework for combining information expressed in different modalities, which suggests an important shared core among the various theories.

Fuzzy sets are a basis for possibility theory. The possibility distribution is a flexible constraint on the value of a variable of interest, which can be interpreted as a fuzzy restriction (Zadeh, 1978). In practical terms, Augustin et al. (2014) consider possibility and its converse, necessity, to be types of upper and lower probabilities. The possibility distribution and measure were originally defined in Zadeh (1978) and Dubois and Prade (1987). Augustin et al. (2014) also discuss the relationship between possibility theory and imprecise probability.

Definition 3. (Zadeh, 1978) A *possibility distribution* is a function $\pi : \Omega \to [0, 1]$ with $\pi(\omega) = 1$ for at least one $\omega \in \Omega$, the outcome space. A *possibility measure*, Π, is a function $\Pi : \wp(\Omega) \to [0, 1]$ defined as $\Pi(X) = \sup_{x \in X} \pi(x)$ with the property that for any $\mathcal{X} \subseteq \wp(\Omega)$, $\Pi\left(\bigcup_{X \in \mathcal{X}} X\right) = \sup_{X \in \mathcal{X}} \Pi(X)$, where X denotes a set of outcomes and $\wp(\Omega)$ denotes the power set of Ω. The *necessity measure* is $N(X) = 1 - \Pi(X^c) = 1 - \sup_{x \in X^c} \pi(x)$.

Dubois et al. (2001) consider the possibility distribution to be a generalized characteristic function of a fuzzy set. The possibility measure is also a particular case of

the fuzzy measure (Zadeh, 1978). Necessity measures are a subset of the functions that can be represented in the Dempster-Shafer theory of evidence (Walley, 2000).

2.2.4 Evidence Theory

Dempster (1967) developed another early generalization of probability theory to upper and lower probabilities, which was formalized and extended by Shafer (1976) into the mathematical theory of evidence. Dempster-Shafer belief and plausibility functions are also types of upper and lower probabilities, similar to possibility and necessity functions. Fine (1977) summarizes the foundation for the theory as follows.

Definition 4. (Fine, 1977) The theory of evidence is based on the frame of discernment, basic probability function, belief function, and plausibility function.

1. The frame of discernment, Ω, is the set of all elements of interest.

2. The basic probability function, $m : \wp(\Omega) \rightarrow [0,1]$, is a function such that $m(\emptyset) = 0$ and $\sum_{X \subset \Omega} m(X) = 1$.

3. The *belief function*, $Bel : \wp(\Omega) \rightarrow [0,1]$, satisfies $Bel(\emptyset) = 0$, $Bel(\Omega) = 1$, and $\forall\, n, Bel\left(\bigcup_{i=1}^{n} X_i\right) \geq \sum_{k=1}^{n} \left(\sum_{i_1 < i_2 < \ldots < i_k} (-1)^{k+1} Bel\left(\bigcap_{j=1}^{k} X_{i_j}\right)\right)$

4. The *plausibility function*, $Pl : \wp(\Omega) \rightarrow [0,1]$, satisfies $Pl(X) = 1 - Bel(X^c)$.

The interaction between the basic definitions and the rules for combining evidence in Dempster-Shafer theory have been criticized from the perspective of more general uncertainty theories (Zadeh, 1984; Walley, 1987). The *transferable belief model* of Smets and Kennes (1994) addresses the core problem by considering "open-world" and "closed-world" contexts that do or do not allow $m(\emptyset) > 0$. This change addresses the situation where events outside the frame of discernment are possible and assigned to the empty set. Smets and Kennes (1994) emphasize that their construction of the

transferable belief model does not require any standard probability theory, although it can be interpreted in a way that is compatible with a probabilistic foundation.

The frame of discernment is equivalent to the sample space or outcome space in other contexts and the last criterion in Definition 4, Item 3 is equivalent to saying the belief function is ∞-monotone. In fact, belief functions are a subset of ∞-monotone capacities, so there is a close connection to fuzzy measure theory.

2.2.5 Fuzzy Measure Theory

Fuzzy measures, also referred to as *capacities*, generate a broad class of uncertainty theories (Choquet, 1954). In particular, fuzzy measures generalize standard measures and probability measures by replacing additivity with monotonicity (Sugeno, 1974). Murofushi and Sugeno (1991) define a fuzzy measure as follows.

Definition 5. (Murofushi and Sugeno, 1991) Given a non-empty set Ω and a σ-algebra $\mathcal{F}$, a *fuzzy measure*, or capacity, is a function $f : \mathcal{F} \to (-\infty, \infty)$ such that

1. For all $E \in \mathcal{F}$, $f(E) \geq 0$
2. $f(\emptyset) = 0$
3. $f(A) \leq f(B)$ whenever $A \subset B$ and $A, B \in \mathcal{F}$

Augustin et al. (2014) highlight several properties related to capacities that are useful for comparing uncertainty theories.

Definition 6. (Augustin et al., 2014) Set operations for a capacity, f, can be classified as super-additive, sub-additive, n-monotone, and ∞-monotone.

1. Super-additive: $A \cap B = \emptyset \Rightarrow f(A \cup B) \geq f(A) + f(B)$.
2. Sub-additive: $A \cap B = \emptyset \Rightarrow f(A \cup B) \leq f(A) + f(B)$.

3. n-monotone: Let $n \in \mathbb{N}_0, n \geq 2$. A capacity f is n-monotone if for any collection $\mathcal{A}_n \subseteq \wp(\Omega)$ of n events, it holds that

$$f\left(\bigcup_{A \in \mathcal{A}_n} A\right) \geq \sum_{\emptyset \neq \mathcal{A}' \subseteq \mathcal{A}_n} (-1)^{|\mathcal{A}'|+1} f\left(\bigcap_{A \in \mathcal{A}'} A\right). \tag{1}$$

4. ∞-monotone: A capacity f is ∞-monotone whenever it is n-monotone for all $n \in \mathbb{N}_0, n \geq 2$.

Capacities, or fuzzy measures, include interval probabilities and lower and upper probabilities as prominent special cases. These types of probabilities are often used in safety and reliability applications in the engineering domain. Coherent lower and upper probabilities correspond to *super-additive* and *sub-additive* capacities (Augustin et al., 2014). For standard probability theory, probability is additive (i.e., it is both super-additive and sub-additive) and ∞-monotone. Note that if the inequality in the n-monotone property is replaced with equality, it becomes the calculation for the probability of the union of n sets under standard probability theory.

2.2.6 Imprecise Probability

Imprecise probability is a unifying framework for a large class of quantitative uncertainty models. The theory of *sets of desirable gambles*, which has a one-to-one correspondence with partial preference orderings, is the most general model of uncertainty (Walley, 2000). Sets of desirable gambles include coherent lower previsions, which are a generalization of lower probabilities, as a prominent special case. Any uncertain phenomenon that can be represented in a more specific model discussed in the previous subsections can be represented by sets of desirable gambles, and some aspects of uncertainty can be represented by sets of desirable gambles that cannot be represented in narrower models.

The core element in imprecise probability theory is the gamble, which can be interpreted as the utility associated with each potential outcome of an experiment or a generic random variable.

Definition 7. (Walley, 2000) A gamble is a bounded mapping $g : \Omega \to \mathbb{R}$, where Ω is the set of possible outcomes under consideration.

A coherent set of desirable gambles can be interpreted as all the gambles that are attractive to a decision maker, along with all the gambles rationally implied by those judgments. Mathematically, it is a convex cone of gambles that contains all positive gambles but not the zero gamble (Walley, 2000).

Definition 8. (Walley, 2000) Let $\mathcal{L}$ denote the set of all gambles. For $g, h \in \mathcal{L}$, let $g \geq h$ mean $g(\omega) \geq h(\omega)$ for some $\omega \in \Omega$, and let $g > h$ mean $g \geq h$ and $g(\omega) > g(\omega)$ for some $\omega \in \Omega$. A *set of desirable gambles*, denoted by $\mathcal{D}$, is a subset of $\mathcal{L}$. A set of desirable gambles is coherent if it satisfies the following axioms:

1. $0 \notin \mathcal{D}$
2. If $g \in \mathcal{L}$ and $g > 0$, then $g \in \mathcal{D}$
3. If $g \in \mathcal{D}$ and $c \in \mathbb{R}_+$, then $cg \in \mathcal{D}$
4. If $g \in \mathcal{D}$ and $h \in \mathcal{D}$, then $g + h \in \mathcal{D}$
5. If $g \in \mathcal{L}, g \neq 0, \mathcal{B}$ is a partition of Ω, and $Bg \in \mathcal{D} \cup \{0\}$ for all $B \in \mathcal{B}$, then $g \in \mathcal{D}$

Sets of desirable gambles include all the previously discussed models of uncertainty as special cases (Walley, 2000). Notably, it is a general formulation of nonadditive probability which includes possibility theory, evidence theory, and fuzzy measure theory as commonly used special cases. However, there are several types of uncertainty

that can be developed without relying on the utility scale used for sets of desirable gambles. Dempster-Shafer theory of evidence, possibility theory, and non-monotonic reasoning can all be developed in alternative ways without relying on a numerical foundation (Augustin et al., 2014).

2.2.7 Summary

The models discussed previously have varying levels of generality and modeling fidelity. In the following discussion, we present a framework that describes the relationship among these uncertainty models.

Augustin et al. (2014) present a summary of the relationship among the various models of uncertainty. The model as depicted in Figure 1 is comprehensive but adopts different terminology than used here. One benefit to this format is that it includes the generalization relationships between classes as well as specific examples within each class. Note that the distinction between a class and an element is somewhat arbitrary and the elements could also be considered classes. For instance, the linear previsions element could be a class that includes both countably-additive and finitely-additive probability.

Using these models as a basis, along with information in Walley (2000) and Destercke et al. (2008), we develop and present a unified framework for uncertainty in Figure 2. The arrow relationship $A \rightarrow B$ indicates that A generalizes B in the sense that models in B are a subset of the models that can be formulated in A. Although this framework suggests a clear distinction between uncertainty and utility, there are connections across classes. This distinction is particularly true for the more general models, but the separation helps distinguish models with different purposes and highlight that models from different classes can be combined.

There is a trade-off between complexity and generality for uncertainty mod-

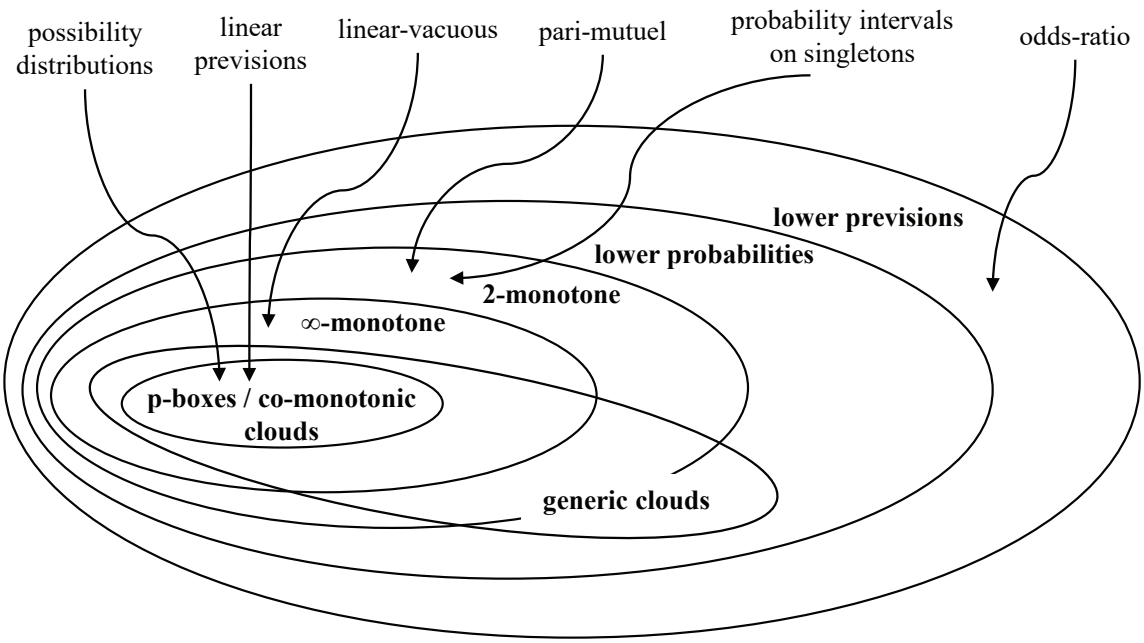

Figure 1. Uncertainty models, adapted from Augustin et al. (2014)

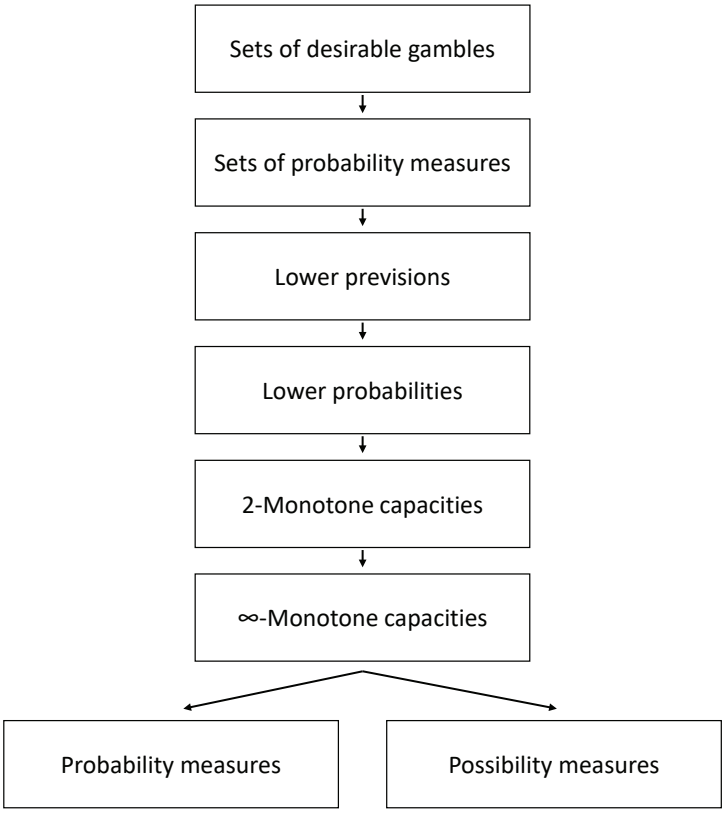

Figure 2. Framework for uncertainty models

els. This trade-off suggests that whenever a problem is simple enough to be well-represented with standard probability, expected utility, and set theory, those are the preferred approaches. If the problem is too complex to be addressed with those standard tools, there is no dominant approach among the remaining uncertainty models. Dempster-Shafer theory is particularly well-known and mature in computer science applications (Smets, 1999) while sets of probability measures have been well integrated into state-of-the-art optimization theory (Wiesemann et al., 2014). Although the choice of uncertainty model is application dependent, standard-form ambiguity sets are highly expressive while remaining computationally tractable. In addition to the models discussed here, there are other uncertainty models including clouds (Neumaier, 2004), pari-mutuel (Walley, 1991), odds-ratio (Berger et al., 1994), and ϵ-contamination (Huber, 1964).

2.3 Decision Making Under Uncertainty

Representations of uncertainty are an important component of decision theory and applied decision making. This section summarizes the relationship between uncertainty models and the classical and contemporary decision theory models. Etner et al. (2012), Starmer (2010), and Marinacci (2015) provide surveys of decision theory under ambiguity, which include further details on economic considerations and experimental evidence. There is a significant body of research that investigates the experimental evidence for a wide range of decision models (Dimmock et al., 2015a,b; Binmore et al., 2012; Baillon et al., 2018). However, this survey focuses on the decision models rather than the empirical findings.

There are several alternative approaches to modeling uncertainty in decision making. This survey briefly discusses qualitative approaches but is primarily focused on quantitative methods. Within quantitative methods, we identify several classes

of related models including expected utility, prospect theory, multiple priors, variational preferences, second-order beliefs, information models, possibilistic and imprecise models, and alternative models. We discuss each class in detail in the following subsections, with an extended presentation of expected utility and prospect theory as the rational and behavioral foundations, respectively, for many subsequent models examined in this dissertation.

2.3.1 Expected Utility

The foundation of expected utility theory is the seminal Von Neumann-Morgenstern (VNM) Expected Utility Theorem, based on four rationality axioms. Expected utility theory primarily develops a theory for utility and decision making, but it also makes assumptions about uncertainty. Following the structure used in Levin (2006), we have the following exposition.

Definition 9. (von Neumann and Morgenstern, 1947) A lottery is a probability distribution, P, over an outcome space, Ω, with $P \in \mathscr{M}$, the set of all possible lotteries. When Ω is a finite set of cardinality n, let $P = (P_1, P_2, \ldots, P_n)$ where $P_i = P(\omega_i)$.

The following rationality axioms are relatively mild, but the completeness axiom does imply a precise and comprehensive ability to compare preferences that restricts expressiveness. VNM rationality is defined by four axioms over lotteries: completeness, transitivity, continuity, and independence.

Definition 10. Let $P \succcurlyeq Q$ indicate that P is weakly preferred to Q. Then,

1. Completeness: For any two lotteries, P and Q, $P \succcurlyeq Q$, $Q \succcurlyeq P$, or $P \sim Q$.

2. Transitivity: For any lotteries P, Q, and R, if $P \succcurlyeq Q$ and $Q \succcurlyeq R$, then $P \succcurlyeq R$.

3. Continuity: For any lotteries $P, Q,$ and R with $P \succcurlyeq Q \succcurlyeq R$, there exists some $\alpha \in [0,1]$ such that $\alpha P + (1-\alpha)R \sim Q$.

4. Independence: For any lotteries $P, Q,$ and R and $\alpha \in [0,1]$, $P \succcurlyeq Q \iff \alpha P + (1-\alpha)R \succcurlyeq \alpha Q + (1-\alpha)R$.

The following definition formalizes the notion of expected utility, assuming a finite outcome space.

Definition 11. (von Neumann and Morgenstern, 1947) A utility function $U : \mathscr{M} \to \mathbb{R}$ has an *expected utility* form if there are numbers $(u_1, u_2, \ldots, u_n)$ for each of the n outcomes in $(\omega_1, \omega_2, \ldots, \omega_n)$ such that for every $P \in \mathscr{M}$, $U(P) = \sum_{i=1}^{n} p_i u_i$.

The key assumption of the VNM expected utility theorem is that a rational decision maker will always act to maximize expected utility.

Theorem 1. *(von Neumann and Morgenstern, 1947) A complete and transitive preference relation $\succcurlyeq$ on $\mathscr{M}$, the set of all possible lotteries, satisfies continuity and independence if and only if it admits an expected utility representation.*

VNM expected utility can be considered the standard for decision theory. *Subjective expected utility* (Savage, 1954) offers a compelling axiomitization using subjective probabilities that also leads to an expected utility maximization framework. Expected utility explicitly builds a decision theoretic foundation on the standard, Kolmogorov probability theory discussed in Section 2.2.1. However, the basic expected utility structure has been extended in a variety of ways by removing or modifying axioms. Machina (1982) presents an expected utility variant without the independence axiom. Removing the completeness axiom results in a rich theory with sets of utility functions (Dubra et al., 2004). The extension to set-valued utility can also be combined with set-valued probability (Weber, 1987).

Superficially, the VNM concept of a lottery and the imprecise probability concept of a gamble are similar, but there is an important, complementary distinction. The lottery is a probably distribution over outcomes, whereas the gamble is a real-valued function over outcomes. VNM theory uses preferences and probability to study utility, whereas imprecise probability theory uses preferences and utility to study probability.

2.3.2 Prospect Theory

Prospect theory is a theory for behavioral decision making under uncertainty (Kahneman and Tversky, 1979). The motivation for prospect theory is that, in practice, decision makers violate rationality assumptions for both utility and probability. These violations are incorporated into *cumulative prospect theory* as a non-linear utility function and a non-linear probability weighting function, which are improvements from the original theory to address violations of stochastic dominance constraints (Tversky and Kahneman, 1992). Formally, Wakker (2010) defines prospect theory using the following definitions. We omit some details for brevity.

First, a *prospect* is a payoff based on the outcome of an uncertain process. The terminology in prospect theory conflicts with the previously discussed models, so we reframe the language in several cases. Given an outcome space, denoted by the set Ω, an event, E, is a subset of the outcome space, $E \in \wp(\Omega)$. Formally, a prospect is a function $r : \wp(\Omega) \to \mathbb{R}$, which yields reward r_1 under event $E_1 \ldots$ and r_n under event E_n and is also denoted as $r = E_1 r_1 \ldots E_n r_n$ (Wakker, 2010).

Next, the preference relation, $\succcurlyeq$, is defined as a natural way of ordering outcomes by desirability. Notably, $\succcurlyeq$ is complete, while several important choice functions in imprecise probability (e.g., E-admissibility) are not. If a representing function exists, then $\succcurlyeq$ is a weak order (Wakker, 2010). That is, $\succcurlyeq$ satisfies:

Transitivity: for prospects r, s, and t if $r \succcurlyeq s$ and $s \succcurlyeq t$ then $r \succcurlyeq t$, and

Completeness: for prospect r and s, $r \succcurlyeq s$ or $s \succcurlyeq r$.

These definitions are used to make the key assumption for prospect theory, in particular the finite nature of the reward space and the complete preference relation.

Assumption 1. *(Wakker, 2010)* Ω *is a, finite or infinite, outcome space, and* $\mathbb{R}$ *is the reward set. Prospects map subsets of outcomes to rewards, taking only finitely many values. The domain of preference is the set of all prospects, i.e., of all such maps.* $\succcurlyeq$ *is a preference relation on the set of prospects. Nondegeneracy holds.*

Given Assumption 1, prospect theory makes two major behavioral adjustments to expected value theory. First, the reward values, $r(E)$, are modified by a general utility function (Wakker, 2010). Second, the probabilities are modified by a specific type of function that is derived from rank dependent utility theory (Quiggin, 1982). Rather than weighting individual probabilities directly, rank dependent utility theory weights probability "ranks," which are the complement of quantiles (Wakker, 2010). Consistent with behavioral experiments, the probability weight functions are allowed to differ for losses and gains, and the weighted probabilities no longer sum to one (Wakker, 2010). Both of these modifications are applied after framing and referencing the problem to a zero utility value (Wakker, 2010).

Definition 12. (Wakker, 2010) *Prospect theory* holds if there exists a strictly increasing continuous utility function $U : \mathbb{R} \to \mathbb{R}$ with $U(0) = 0$ and two weighting functions W^+ and W^-, denoted as π, such that each prospect $r = E_1 r_1 \cdots E_n r_n$ with completely sign-ranked rewards $r_1 \geq \cdots \geq r_k \geq 0 \geq r_{k+1} \geq \cdots \geq r_n$ for some $0 \leq k \leq n$ is evaluated by

$$\begin{aligned} \sum_{j=1}^{n} \pi_j U(r_j) &= \sum_{i=1}^{k} \pi(E_i^{E_{i-1} \cup \cdots \cup E_1}) U(r_i) + \sum_{j=k+1}^{n} \pi(E_{j_{E_{j+1} \cup \cdots \cup E_n}}) U(r_j) \\ &= \sum_{i=1}^{k} (W^+(E_i \cup \cdots \cup E_1) - W^+(E_{i-1} \cup \cdots \cup E_1)) U(r_i) \end{aligned} \tag{2}$$

$$+\sum_{j=1}^{k}(W^{-}(E_j \cup \cdots \cup E_n) - W^{-}(E_{j+1} \cup \cdots \cup E_n))U(r_j), \qquad (3)$$

the prospect theory value of the prospect.

The aspects of prospect theory relating directly to uncertainty have been formalized under the name *support theory* (Tversky and Koehler, 1994). Support theory is a version of sub-additive probability that models behavioral attitudes toward risk (Tversky and Koehler, 1994). Tversky and Koehler (1994) provide experimental evidence for descriptive behavior compatible with support theory. The theoretical structure of the sub-additive probability is not significantly different from various other types of lower probabilities, which are a particular case of fuzzy measures discussed in Section 2.2.5.

Prospect theory is closely associated with *rank dependent utility* (Quiggin, 1982; Schmeidler, 1989), a particular form of the general nonadditive probability models discussed in Section 2.2.5. However, variants of prospect theory using sets of probability distributions, interval probabilities, and other uncertainty models have also been developed (Wakker, 2010; Diecidue and Wakker, 2001). These variants combine aspects of the decision models in the following subsections with the prospect theory treatment of utility functions.

Prospects in prospect theory and gambles in imprecise probability, discussed in Section 2.2.6, are nearly equivalent notions, but prospects are defined over events (i.e., sets of outcomes) (Wakker, 2010) whereas gambles are defined directly over outcomes. Recent work has extended prospects from finitely many values to infinitely many values (Kothiyal et al., 2011). However, prospect theory focuses on behavioral decision making, whereas imprecise probability focuses on modeling uncertainty. The main area of overlap between prospect theory and imprecise probability is with respect to elicitation. The elicitation of imprecise probabilities is a relatively under-developed

field (Augustin et al., 2014), which may benefit from more explicit connection to behavioral results from prospect theory.

2.3.3 Multiple Priors

Decision makers often lack sufficient information to credibly define a unique prior distribution. In the presence of this uncertainty, one solution is to base the decision making model on multiple priors modeled as a set of probability distributions, which is a flexible uncertainty model discussed in Section 2.2.2. By replacing the single distribution with a set of distributions, the decision maker can represent the ambiguity about the true prior distribution.

The main multiple priors models directly generalize the classic *maxmin utility* and α *maxmin utility* approaches to sets of probability distributions. In the classic Wald (1945) maxmin model, the decision maker maximizes utility with respect to the worst possible outcome, regardless of likelihood. The α maxmin model, due to Arrow and Hurwicz (1972), generalizes the maxmin criterion by considering a convex combination of the worst outcome and the best outcome, $\alpha\underline{u} + (1-\alpha)\bar{u}$. The best and worst outcomes are represented by $\bar{u}$ and $\underline{u}$, respectively, and the α coefficient represents the decision maker's pessimism.

In the multiple priors generalization, the maxmin criterion is replaced with *maxmin expected utility* in Gilboa and Schmeidler (1989). Let Ω be the outcome space representing the possible states of the world, $\mathcal{C}$ be the consequence space representing any abstract consequence such as money or health, and $f : \Omega \to \mathcal{C}$ be a decision or act representing a course of action. Then the decision, f, is evaluated by

$$\min_{\mathbb{P}\in\mathcal{M}} \mathbb{E}_{\mathbb{P}} u(f), \tag{4}$$

where $\mathbb{P}$ is a probability distribution over outcomes in Ω and $\mathbb{P}$ is included in the set

of priors, $\mathcal{M}$. The expected utility operator with respect to probability distribution $\mathbb{P}$ is denoted by $\mathbb{E}_{\mathbb{P}}$. The utility function, $u(f) : \mathcal{C} \rightarrow \mathbb{R}$, represents the utility of decision f by defining a real-valued utility for the consequence associated with any outcome through f. In this maxmin expected utility model, the ambiguity averse (i.e., uncertainty averse) decision maker maximizes expected utility according to the least favorable probability distribution in the set of prior distributions. Similarly, Ghirardato et al. (2004) generalize the α maxmin approach to multiple priors in α *maxmin expected utility*. In this model, the decision f is evaluated by

$$\alpha \min_{\mathbb{P} \in \mathcal{M}} \mathbb{E}_{\mathbb{P}} u(f) + (1 - \alpha) \max_{\mathbb{P} \in \mathcal{M}} \mathbb{E}_{\mathbb{P}} u(f), \tag{5}$$

where α represents the ambiguity aversion of the decision maker. Although the α parameter allows the decision maker to avoid the worst-case approach of maxmin expected utility, the α parameter does not have a clear interpretation. In particular, $\alpha = 0.5$ does not correspond to expected utility maximization, and the ambiguity aversion associated with α cannot be cleanly separated from the ambiguity aversion associated with the size of the set of distributions (Etner et al., 2012).

Choquet expected utility is also closely related to this family of models, but is more general in the sense that it does not assume ambiguity aversion. This model, due to Schmeidler (1989), evaluates a decision f by

$$\int u(f) d\nu, \tag{6}$$

where $\int$ is the Choquet integral and ν is a capacity. A capacity generalizes a probability function and is not necessarily additive, as discussed in Section 2.2.5. The Choquet integral is necessary to address this special structure. Although the Choquet expected utility model allows for non-ambiguity averse decision makers, not all

sets of probability distributions can be fully represented as a capacity. Chateauneuf et al. (2007) present an extension of Choquet expected utility to a Hurwicz-type optimism-pessimism framework with neo-additive capacities, which we refer to as *neo-Choquet expected utility*.

2.3.4 Variational Preferences

Variational preferences are another way of generalizing maxmin expected utility to non-ambiguity averse decision making, developed in Maccheroni et al. (2006). The decision maker evaluates decision f by

$$\min_{\mathbb{P}\in\mathcal{P}} \left[\mathbb{E}_{\mathbb{P}} u(f) + c(\mathbb{P})\right] = \min_{\mathbb{P}\in\mathcal{P}} \left[\left(\int u(f) d\mathbb{P}\right) + c(\mathbb{P})\right], \tag{7}$$

where $\mathcal{P}$ is the set of all probability distributions on the state space and $c(\mathbb{P})$ is a cost function. This decision model also uses sets of probability measures as the foundational uncertainty model, as discussed in Section 2.2.2. Several notable decision models correspond to specific forms of the cost function (Etner et al., 2012). The different models are summarized in Table 1, in which the notional or hypothesized prior probability distribution is denoted by $\mathbb{Q}$.

The *ε-contamination* model (Eichberger and Kelsey, 1999) also assumes the decision maker has a suspected prior distribution, but the decision maker is only $(1-\varepsilon)\%$ confident the distribution is correct. To address the uncertainty, the decision maker considers a combination of the suspected distribution, $\mathbb{Q}$, and any arbitrary distribution, $\mathcal{P}$, $(1-\varepsilon)\mathbb{Q} + \varepsilon\mathcal{P}$. Note that this is a special case of the maxmin expected utility approach where the set of priors, $\mathcal{M}$, has the specified ε-contamination form. The cost function then acts as an indicator for membership in the set of priors.

The *robust control* model (Hansen and Sargent, 2001), or *multiplier preferences* model, assumes the decision maker has a specific prior, $\mathbb{Q}$, that is the default prior.

Table 1. Variational preference cost functions

Cost Function	Corresponding Model
$c(\mathbb{P}) = \begin{cases} 0 & \text{if } \mathbb{P} = \mathbb{Q} \\ \infty & \text{otherwise} \end{cases}$	Expected utility
$c(\mathbb{P}) = 0$	Maxmin utility
$c(\mathbb{P}) = \begin{cases} 0 & \text{if } \mathbb{P} \in \mathcal{M} \\ \infty & \text{otherwise} \end{cases}$	Maxmin expected utility
$c(\mathbb{P}) = \begin{cases} 0 & \text{if } \mathbb{P} \in (1-\varepsilon)\mathbb{Q} + \varepsilon\mathcal{P} \\ \infty & \text{otherwise} \end{cases}$	ε-contamination
$c(\mathbb{P}) = \theta R(\mathbb{P} \,\|\|\, \mathbb{Q})$	Robust control

To account for uncertainty about this prior, the cost function is defined using the Kullback-Leibler divergence, or relative entropy, $R(\mathbb{P} \,||\, \mathbb{Q})$. The relative entropy cost function weights probability distributions that are closer to the suspected prior higher than probability distributions that are dissimilar. The parameter θ controls the strength of this weighting and indirectly represents the decision maker's ambiguity aversion. Unlike the previous models, this approach allows smooth weighting over the set of potential prior distributions.

2.3.5 Second-order Beliefs

Klibanoff et al. (2005) present a *smooth* model of decision making under uncertainty based on second-order beliefs. In this model, the decision maker has first-order beliefs about the likelihood of states of the world, represented as a probability distribution, as in decision making under risk. However, the decision maker also has second-order beliefs about the likelihood each probability distribution is correct, represented as a probability distribution over probability distributions. What dis-

tinguishes this model from a second-order Bayesian model that reduces to expected utility is that Klibanoff et al. (2005) also modify the inner expectation by a function, Φ, that controls ambiguity aversion. They evaluate a decision f by

$$\mathbb{E}_{\mathbb{L}\in\mathcal{L}}\Phi(\mathbb{E}_{\mathbb{P}\in\mathcal{P}}(u(f))), \tag{8}$$

where $\mathbb{L}$ is a probability distribution in the set of probability distributions, $\mathcal{L}$, over the inner probability distributions. This model for uncertainty can be interpreted as a generalization of sets of probability measures, discussed in Section 2.2.2. However, this approach is distinct because of the second-order probability distribution $\mathbb{L}$. While the standard set of probability measures has no probabilistic information over the first-order distributions, the uncertainty model underlying smooth ambiguity introduces a family of second-order distributions over those first-order distributions. Notably, without the Φ function, this uncertainty formulation would collapse into a standard set of uncertainty measures using Bayesian conditioning. The Φ function controls the ambiguity aversion in the same way the utility function is related to risk aversion. A concave Φ corresponds to an ambiguity averse decision maker and can represent maxmin expected utility in the limiting case of infinite ambiguity aversion.

Second-order subjective expected utility (Seo, 2009) with generalizations by Nascimento and Riella (2013), *state-dependent* models (Nau, 2006), and *second-order probabilistic sophistication* (Ergin and Gul, 2009) provide alternative axiomatizations of second-order beliefs that result in a similar decision criterion to the criterion defined by Klibanoff et al. (2005). Klibanoff et al. (2009) also extend the static smooth model to the dynamic case. Lang (2017) explores the differences in behavior resulting from different approaches to second-order belief models.

Cerreia-Vioglio et al. (2011) present *uncertainty averse preferences* as a generalization that includes variational preferences and the smooth model with concave Φ

as special cases. In this model a decision f is evaluated by

$$\min_{\mathbb{P}\in\mathcal{P}} G\left(\mathbb{E}_{\mathbb{P}} u(f), \mathbb{P}\right) = \min_{\mathbb{P}\in\mathcal{P}} G\left(\int u(f) d\mathbb{P}, \mathbb{P}\right) \tag{9}$$

where $G : u(X) \times \mathcal{P} \to (-\infty, \infty]$ is a quasiconvex function controlling ambiguity attitude and X is the outcome space. If $G(t, \mathbb{P}) = t + c(\mathbb{P})$, the model corresponds to variational preferences (Maccheroni et al., 2006). If $G(t, \mathbb{P}) = t + \min_{\nu\in\Gamma(\mathbb{P})} I_t(\nu \,||\, \mu)$, the model corresponds to the smooth model (Klibanoff et al., 2005). $\Gamma(\mathbb{P})$ is the set of all second-order probabilities, ν, that are absolutely continuous with respect to a prior, μ, and have $\mathbb{P}$ as their reduced probability measures over the state space, S. $I_t(\nu \,||\, \mu)$ is a statistical distance function to the prior distribution. Unlike the smooth ambiguity decision model, the uncertainty model associated with this decision model is a standard set of probability distributions, as discussed in Section 2.2.2.

2.3.6 Information Models

The subjective process for identifying the set of prior distributions in multiple prior models is not addressed in the previous formulations. Gajdos et al. (2008) present a decision model, sometimes called the *contraction* model, that explicitly includes the objective information available to the decision maker. Gajdos et al. (2008) represent this information as a set of probability distributions, $\mathcal{Q}$, and introduce a subjective mapping, φ, from this information to the set of priors, $\mathcal{M}$, such that $\varphi(\mathcal{Q}) = \mathcal{M}$. The decision maker then uses this set of priors to make decisions as in the maxmin expected utility framework. A decision f is evaluated by

$$\min_{\mathbb{P}\in\varphi(\mathcal{Q})} \mathbb{E}_{\mathbb{P}} u(f). \tag{10}$$

Under additional axioms, Gajdos et al. (2008) present a specific functional form for this evaluaiton as

$$\alpha \min_{\mathbb{P}\in\mathcal{Q}} \mathbb{E}_{\mathbb{P}} u(f) + (1-\alpha)\mathbb{E}_{s(\mathcal{P})} u(f), \tag{11}$$

where $s(\mathcal{P})$ is the Steiner point of the set of distributions, analogous to a mean value. The parameter α represents the ambiguity aversion. When $\alpha = 0$, the model corresponds to expected utility and when $\alpha = 1$, the model corresponds to maxmin expected utility. The uncertainty model supporting these decision models is sets of probability distributions, as discussed in Section 2.2.2.

Jaffray (1989) also considers the objective information available to the decision maker in *linear utility for belief functions*, but does not address subjective formation of a set of priors. Instead, he works with capacities in the imprecise probability approach. The capacity associated with decision f, ν_f ,is evaluated as

$$\sum_{E\in\mathcal{A}} \varphi(E) \left[\alpha(m_E, M_E)u(m_E) + (1-\alpha(m_E, M_E))u(M_E)\right], \tag{12}$$

where E is an event in the set of all events $\mathcal{A}$ and m_E and M_E are the minimal and maximal outcomes on event E, respectively. The $\alpha(m_E, M_E)$ function controls the pessimism of the decision maker with respect to uncertainty. The function φ is the Mobius transform of the capacity ν_f, which is associated with the ambiguity of an event in terms of lower probabilities. If the α function is constant, this model is equivalent to α maxmin expected utility with an objective set of prior distributions (Etner et al., 2012). In the general case, the underlying uncertainty model is nonadditive capacities, as discussed in Section 2.2.5.

2.3.7 Possibilistic and Imprecise Models

Whalen (1984) presents a generalization of the Wald maxmin criterion using possibility theory as the uncertainty model. In this decision model, a decision f is evaluated by

$$\inf_{s \in S} \max n(\pi(s)), \mu(f(s)) \tag{13}$$

where s is a state in the set of states, S, π is a normalized possibility distribution with values in a plausibility scale, μ is a possibility distribution with values on a preference scale, and n is an order-reversing map which is $n(x) = 1 - x$ when applied to normalized arguments. Yager (1979) presents an optimistic counterpart to this criterion, and Dubois and Prade (2001) develop a decision-theoretic axiomatization of both models which is generalized by Giang and Shenoy (2005) with an alternative axiomitization. These models explicitly use possibility theory as the uncertainty model, as discussed in Section 2.2.3.

In addition to the purely possibilistic models, there are several closely related decision theories. The restriction to Wald-type extreme optimism and pessimism in the previous models is removed in single-stage and multi-stage *one-shot decision theory* (Guo, 2011; Guo and Li, 2014). The one-shot model introduces 12 focus points that characterize the decision maker's attitude about possibility and satisfaction and an optimality criterion which is defined for each focus point. This model explicitly uses possibility theory, discussed in Section 2.2.3, as the underlying uncertainty model. One-shot decision theory is generalized to the *focus theory of choice* in Guo (2019) using relative likelihood and a positive and negative evaluation system. The relative likelihood uncertainty model is a tailored nonadditive model that can be used to scale probabilities by the highest probability event in a subset of events. General nonadditive models are discussed in Section 2.2.5, but the relative likelihood model is distinct because it accounts for salience and decision maker behavior. The focus

theory of choice accounts for the behavioral distinction between positive and negative frames, salience of payoffs, and provides a formal model for procedural rationality that can be used as a theoretical foundation for behavioral decision models.

Fuzzy, interval, and imprecise probabilities have also been considered in the decision theory literature. Bellman and Zadeh (1970) formulate a structure for decision making in fuzzy environments. Guo and Tanaka (2010) present decision criteria for dominance, indifference, and incomparable alternatives using interval expected values. Troffaes (2007) also extends this line of work to models using imprecise probability as the uncertainty foundation. He introduces several alternative criteria including admissibility, maximal expected utility, maximality, E-admissibility, and maxmin expected utility. The interval probability and imprecise uncertainty models used in these decision models are discussed in Section 2.2.5 and Section 2.2.6.

2.3.8 Alternative Models

In addition to the classes of decision models discussed previously, we briefly highlight several alternative models. Modifications to traditional rationality axioms include regret with non-transitive preferences (Loomes and Sugden, 1982), incomplete preferences (Eliaz and Ok, 2006), general bounded rationality (Conlisk, 1996), and subjective weighting (Hazen, 1989). Another stream of research considers stochastic preferences in the trembling hand model (Harless and Camerer, 1994), random preferences (Loomes and Sugden, 1995), and random expected utility (Gul and Pesendorfer, 2006).

Recent advances at the intersection of neuroscience and economics have led to an active area of research. Fehr and Rangel (2011) provide a recent review of neuroeconomic models. Krajbich et al. (2014) define a drift-diffusion model of choice based on physical neurological processes. The stochastic and neuroeconomic sub-fields have

also been combined in recent work (Woodford, 2014; Natenzon, 2019).

Some decision models take a qualitative perspective on uncertainty, as in the deep uncertainty literature (Walker et al., 2003; Lempert, 2003; Ben-Haim, 2006; Cox, 2012). Deep uncertainty is intended for decision problems for which even a mild quantification, such as an unrestricted uncertainty set, cannot be credibly defined. In many cases, a quantifiable approximation can still provide insight to the decision maker. Learning models, such as those developed by Epstein and Schneider (2007), also help to address situations when the decision maker cannot initially formulate the problem but can learn over time.

2.3.9 Summary

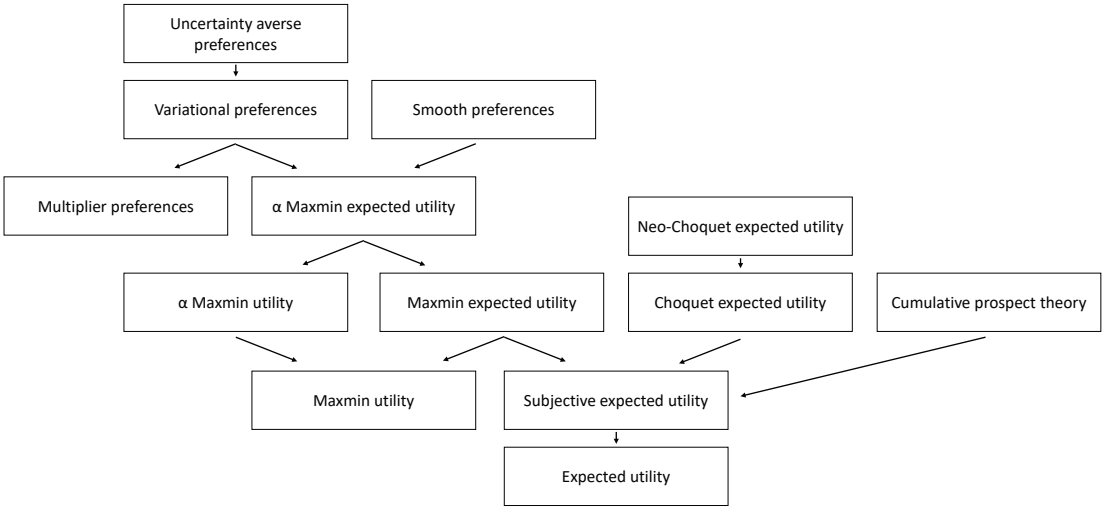

Figure 3. Framework for decision models

Figure 3 summarizes the main relationships among the major categories of decision making models. The arrow relationship $A \to B$ indicates that A generalizes B in the sense that models in B can be formulated as a special case of the more general models in A. Recent results have explored these relationships and introduced successively more generalized decision models. However, the majority of modern

decision-theoretic models use sets of probability distributions as the underlying uncertainty model. Although there are many different approaches to using these sets of probability distributions, the foundations are similar. Generalized models such as variational preferences and smooth preferences emphasize this underlying similarity through their ability to express a variety of more specific models as special cases.

However, there are some notable exceptions to the trend of using sets of probability distributions as the foundation for decision modeling. Some recent work uses capacities rather than sets of probability distributions, such as Choquet expected utility and recent extensions. Other experimentally-oriented theories focus on the specific form and parameters of the decision model to address risk and ambiguity attitudes found in practice. Recent work in this area has focused on smooth models that allow for a distinction between ambiguity level and ambiguity attitude while allowing for arbitrary risk and ambiguity attitudes.

2.4 Optimization Under Uncertainty

Uncertainty models and decision models have been incorporated into optimization using several different approaches. Lim et al. (2006) identify three broad categories of uncertainty models in the operations research literature: uncertainty for variables, distribution parameters, or distributions themselves. The first category corresponds to uncertainty sets and the last two correspond to ambiguity sets. Lim et al. (2006) also highlight that rewards can be standard or benchmarked through regret or a competitive ratio, but all of these reward structures are compatible with any of the uncertainty models. The optimization literature generally focuses on worst-case scenarios in the robust framework, as in maxmin utility decision making, but other decision criteria have become more prevalent in recent years.

Optimization operationalizes decision theory by addressing realistic, large-scale

decision problems. The three main classes of optimization under uncertainty are stochastic, robust, and distributionally robust optimization. We also briefly consider alternative approaches based on non-probabilistic foundations. We present the generic optimization formulation in each case to facilitate comparisons, but much of the literature is also devoted to exploring important sub-classes of general static formulations including linear, quadratic, convex, non-convex, and discrete optimization. These optimization frameworks have also been extended to sequential and adversarial settings. We consider each major class of optimization under uncertainty in detail in the following subsections.

2.4.1 Stochastic Programming

Stochastic programming (Dantzig, 1955; Beale, 1955) models uncertainty using a probability distribution for the uncertain variables. The general stochastic programming problem is $\min_{x \in X} \mathbb{E}_{\mathbb{Q}} f(\mathbf{x}, \omega)$ (Kataoka, 2016; Dupacova et al., 2003). In the following formulation, we formulate the generic stochastic programming problem as a probability constraint to facilitate comparison with robust formulations. Note that an uncertain objective function is a specific case of the constraint formulation if a dummy variable t is introduced such that $f(\mathbf{x}) = t$ and the corresponding constraint is $\mathbb{E}_{\mathbb{Q}}[g(\mathbf{x}, \omega)] \leq t$. Then, the stochastic program is

$$\begin{aligned} \underset{\mathbf{x} \in X}{\text{minimize}} \quad & f(\mathbf{x}) && \text{(14a)} \\ \text{subject to} \quad & \mathbb{E}_{\mathbb{Q}_i}[g_i(\mathbf{x}, \omega_i)] \leq 0 \quad \forall\, i = 1, \ldots, m. && \text{(14b)} \end{aligned}$$

In this model, $\mathbf{x} \in X \subseteq \mathbb{R}^n$ is the decision vector, $f, g_i : \mathbb{R}^n \to \mathbb{R}$ are functions, and ω_i is the stochastic parameter which is distributed according to the probability distribution $\mathbb{Q}_i$. This optimization framework is based on a traditional approach

to uncertainty. It uses a standard probability distribution and a standard expected utility decision criterion, as discussed in Section 2.2.1 and Section 2.3.1, respectively.

While stochastic programming addresses risk, the standard formulation has limited tractability and does not address ambiguity. Such tractability concerns motivate the robust and distributionally robust variants of stochastic programming, which adopt a worst-case approach but achieve significant computational benefits while addressing ambiguity.

2.4.2 Robust Optimization

Robust optimization (Soyster, 1973; Bertsimas and Sim, 2004; Ben-Tal et al., 2009) models uncertainty using uncertainty sets for uncertain variables. Following Bertsimas et al. (2011), the robust optimization formulation is

$$\underset{\mathbf{x} \in X}{\text{minimize}} \quad f(\mathbf{x}) \tag{15a}$$

$$\text{subject to} \quad g_i(\mathbf{x}, \mathbf{u}_i) \leq 0 \quad \forall\, \mathbf{u}_i \in \mathcal{U}_i,\, i = 1, \ldots, m. \tag{15b}$$

In this model, $\mathbf{u}_i \in \mathbb{R}^k$ is the uncertainty parameter, and $\mathcal{U}_i \subseteq \mathbb{R}^k$ is the uncertainty set for some integer k. The robust formulation differs substantially from the stochastic formulation in that there is no probabilistic information associated with the uncertainty sets. This formulation is computationally tractable and supports models where the decision maker cannot credibly identify probability distributions.

The underlying uncertainty model in robust optimization is the uncertainty set, discussed in Section 2.2.2. The underlying decision theoretic model in robust optimization is maxmin utility, discussed in Section 2.3.3. The structure of the uncertainty set in the robust formulation has important implications for tractability. Goerigk and Schöbel (2016) summarize six common uncertainty sets in the recent lit-

erature on robust optimization. We present a condensed list with several additional sets for completeness. In this family of models, the uncertainty is represented by a vector that belongs to an uncertainty set, $\xi \in \mathcal{U} \subseteq \mathbb{R}^M$. We denote the convex hull by $\text{conv}\{\xi^1, \ldots, \xi^N\} = \{\sum_{i=1}^N \lambda_i \xi^i : \sum_{i=1}^N \lambda_i = 1, \lambda \in \mathbb{R}_+^N\}$ and take $\mathbf{A}$ to be a constant matrix. The following list provides a brief description for each type of uncertainty set, but formal definitions are available in works by Bertsimas and Brown (2009), Ben-Tal et al. (2009), and Natarajan et al. (2009). The uncertainty sets are

1. Deterministic: $\mathcal{U} = \{\xi\}$
2. Finite: $\mathcal{U} = \{\xi^1, \ldots, \xi^N\}$
3. Interval-based: $\mathcal{U} = [\underline{\xi}_1, \overline{\xi}_1] \times \cdots \times [\underline{\xi}_M, \overline{\xi}_M]$
4. Polytopic: $\mathcal{U} = \text{conv}\{\xi^1, \ldots, \xi^N\}$
5. Ellipsoidal: $\mathcal{U} = \{\xi \in \mathbb{R}^M : \sqrt{\sum_{i=1}^M \xi_i^2/\sigma_i^2} \leq \Omega\}$ for a parameter $\Omega \geq 0$
6. Norm-based: $\mathcal{U} = \{\xi \in \mathbb{R}^M : \|\xi - \hat{\xi}\| \leq \alpha\}$ for a parameter $\alpha \geq 0$
7. Risk measure-based: $\mathcal{U} = \text{conv}(\{\mathbf{A}\xi : \xi \in \mathcal{X}\})$ for some $\mathcal{X} \subseteq \mathbb{R}^M$
8. Convex: $\mathcal{U} = \text{conv}\{\mathcal{X}\}$ for some $\mathcal{X} \subseteq \mathbb{R}^M$
9. General: $\mathcal{U} = \mathcal{X}$ for some $\mathcal{X} \subseteq \mathbb{R}^M$.

There is a close connection between robust optimization over uncertainty sets and risk measures in mathematical finance. Risk measures can be derived from uncertainty sets (Natarajan et al., 2009) and uncertainty sets can be constructed from risk measures (Bertsimas and Brown, 2009). These connections extend to the multivariate setting, as developed by Bazovkin and Mosler (2015) who extend linear programming with spectral risk measures to the multivariate case. There is also a connection between robust optimization and multi-objective optimization, as discussed by Perny

et al. (2006). A comprehensive survey of multiple criteria decision making and multi-attribute utility theory is available, which includes connections to optimization under uncertainty (Wallenius et al., 2008).

2.4.3 Distributionally Robust Optimization

While robust optimization is a tractable approach for addressing uncertainty, it can be overly pessimistic. Recent work in *distributionally robust optimization* addresses this shortcoming by providing a middle ground between fully specified probabilistic models in stochastic programming and completely non-probabilistic uncertainty sets in robust optimization. Wiesemann et al. (2014) formalize the initial work on distributionally robust optimization. Following the notation from the previous sections, we can express the distributionally robust optimization formulation as

$$\underset{\mathbf{x} \in X}{\text{minimize}} \quad f(\mathbf{x}) \tag{16a}$$

$$\text{subject to} \quad \mathbb{E}_{\mathbb{P}_i}[g_i(\mathbf{x}, \mathbf{z}_i)] \leq 0 \quad \forall\, \mathbb{P}_i \in \mathcal{P}_i,\, i = 1, \ldots, m. \tag{16b}$$

In this model, $\mathbf{z}_i$ is distributed according to $\mathbb{P} \in \Delta^k$ which is a probability distribution in the ambiguity set of probability distributions, $\mathcal{P} \subseteq \Delta^k$. The full k-dimensional probability simplex is represented by Δ^k. The distributionally robust formulation is directly related to the maxmin expected utility (Gilboa and Schmeidler, 1989) and ambiguous stochastic programming (Wozabal, 2012). Using standard techniques with a dummy variable in the objective and the distributionally robust constraint, we recover the maxmin expected utility criterion, $\min_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{u})$. Bertsimas and Brown (2009) also identify an equivalence between uncertainty sets in robust optimization and coherent risk measures, which are closely related to Choquet integrals. Fur-

thermore, Ben-Tal et al. (2010) propose *soft robust optimization*, with generalized constraints of the form

$$\inf_{\mathbb{Q}\in\mathcal{Q}(\epsilon)} \mathbb{E}_{\mathbb{Q}} f(\mathbf{x},\xi) \geq -\epsilon \quad \forall \epsilon \geq 0, \tag{17}$$

where $\{\mathcal{Q}(\epsilon)\}_{\epsilon\geq 0}$ is a set of sets of distributions. This formulation can be shown to be equivalent to robust control (Hansen and Sargent, 2001) and variational preferences (Maccheroni et al., 2006) for certain choices of $\mathcal{Q}(\epsilon)$. These close connections between decision theory under ambiguity and optimization under ambiguity highlight the potential for advances in optimization using recent research in decision theory.

The underlying uncertainty model in distributionally robust optimization is ambiguity sets, which are concrete forms of sets of probability measures, discussed in Section 2.2.2. The underlying decision theoretical model for distributionally robust optimization is maxmin expected utility, as discussed in Section 2.3.3 on multiple priors. As in robust optimization, the structure of the ambiguity set is an important consideration. The following list provides a brief introduction to ten common ambiguity sets, but formal definitions are available in the works by Wiesemann et al. (2014) and Augustin et al. (2014). Using the same notation as in Section 2.2.2, the ambiguity sets are

1. Deterministic: $\mathcal{M} = \{\mathbb{P}_d\}$ where $F_{\mathbb{P}_d}(x,k) = \{1, \text{if } x \geq k; 0, \text{otherwise}\}$ for some k

2. Stochastic: $\mathcal{M} = \{\mathbb{P}_0\}$ for some $\mathbb{P}_0 \in \mathcal{P}$

3. Finite: $\mathcal{M} = \{\mathbb{P}_1, \ldots, \mathbb{P}_N\}$

4. P-box: $\mathcal{M} = \{\mathbb{P} \in \mathcal{P} : F_{\underline{\mathbb{P}}}(x) \leq F_{\mathbb{P}}(x) \leq F_{\overline{\mathbb{P}}}(x) \quad \forall\, x \in \mathbb{R}\}$ for some $\underline{\mathbb{P}}, \overline{\mathbb{P}} \in \mathcal{P}$ subject to mild conditions

5. Uncertainty set: $\mathcal{M} = \{\mathbb{P} \in \mathcal{P} : \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{C}] = 1\}$ for some set $\mathcal{C}$

6. Standard form: $\mathcal{M} = \{\mathbb{P} \in \mathcal{P} : \mathbb{E}_{\mathbb{P}}[\boldsymbol{A}\tilde{\boldsymbol{z}} + \boldsymbol{B}\tilde{\boldsymbol{u}}] = \boldsymbol{b}, \mathbb{P}[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{C}_i] \in [\underline{\boldsymbol{p}}_i, \overline{\boldsymbol{p}}_i] \, \forall i \in \mathcal{I}\}$ subject to mild conditions

7. Polytopic: $\mathcal{M} = \text{conv}(\{\mathbb{P}_1, \ldots, \mathbb{P}_N\})$

8. Norm-based: $\mathcal{M} = \{\mathbb{P} \in \mathcal{P} : \|\mathbb{P} - \hat{\mathbb{P}}\| \leq \alpha\}$ for a parameter $\alpha \geq 0$

9. Convex: $\mathcal{M} = \text{conv}(\mathcal{Q})$ for some $\mathcal{Q} \subseteq \mathcal{P}$

10. General: $\mathcal{M} = \mathcal{Q}$ for some $\mathcal{Q} \subseteq \mathcal{P}$.

We now discuss the "standard form" ambiguity set in some detail, but refer the reader to the work by Wiesemann et al. (2014) for the full definition. This ambiguity set is of particular interest in this survey due to its high level of expressiveness while maintaining tractability. Most optimization models used in the applications considered in this survey use explicit standard form ambiguity or uncertainty models that can also be represented as standard form ambiguity. The standard form ambiguity set is defined as the set of all distributions with conic representable confidence sets subject to two mild conditions and with mean values on an affine manifold (Wiesemann et al., 2014). Mathematically, Wiesemann et al. (2014) present the standard form ambiguity set as

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \begin{array}{ll} \mathbb{E}_{\mathbb{P}}[\boldsymbol{A}\tilde{\boldsymbol{z}} + \boldsymbol{B}\tilde{\boldsymbol{u}}] = \boldsymbol{b}, & \\ \mathbb{P}[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{C}_i] \in \left[\underline{\boldsymbol{p}}_i, \overline{\boldsymbol{p}}_i\right] & \forall i \in \mathcal{I} \end{array} \right\}, \tag{18}$$

where $\mathbb{P}$ represents a joint probability distribution of the random vector $\tilde{\boldsymbol{z}} \in \mathbb{R}^P$ and some auxiliary random vector $\tilde{\boldsymbol{u}} \in \mathbb{R}^Q$, with $\boldsymbol{A} \in \mathbb{R}^{K \times P}$, $\boldsymbol{B} \in \mathbb{R}^{K \times Q}$, $\boldsymbol{b} \in \mathbb{R}^K$, and

$\mathcal{I} = \{1, \ldots, I\}$. The confidence sets $\mathcal{C}_i$ are defined as

$$\mathcal{C}_i = \{(\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \boldsymbol{C}_i \boldsymbol{z} + \boldsymbol{D}_i \boldsymbol{u} \preccurlyeq_{\mathcal{K}_i} \boldsymbol{c}_i\} \tag{19}$$

where $\boldsymbol{C}_i \in \mathbb{R}^{L_i \times P}$, $\boldsymbol{D}_i \in \mathbb{R}^{L_i \times Q}$, $\boldsymbol{c}_i \in \mathbb{R}^{L_i}$, $\underline{\boldsymbol{p}}_i, \overline{\boldsymbol{p}}_i \in [0,1]$, $\underline{\boldsymbol{p}}_i \leq \overline{\boldsymbol{p}}_i \, \forall \, i \in \mathcal{I}$, and the $\mathcal{K}_i$ are proper cones. The standard form ambiguity set is also subject to the following two conditions

(**C1**) The confidence set $\mathcal{C}_I$ is bounded and has probability one, that is, $\underline{\boldsymbol{p}}_I = \overline{\boldsymbol{p}}_I = 1$.

(**C2**) There is a distribution $\mathbb{P} \in \mathcal{P}$ such that $\mathbb{P}[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{C}_i] \in (\underline{\boldsymbol{p}}_i, \overline{\boldsymbol{p}}_i)$ whenever $\underline{\boldsymbol{p}}_i \leq \overline{\boldsymbol{p}}_i, i \in \mathcal{I}$.

Such standard form ambiguity sets are capable of representing ambiguity sets based on higher-order moments, the marginal median, variability measures based on the mean absolute deviation and the Huber loss function (Wiesemann et al., 2014). However, this standard form cannot represent ambiguity sets derived from infinitely many moment restrictions (Wiesemann et al., 2014), which are needed to describe symmetry, independence, or unimodality (Hu and Hong, 2012; Popescu, 2005).

The main benefit to this somewhat involved standard form is that it results in tractable optimization problems given mild restrictions on constraints, see in particular Appendix E of the formal development by Wiesemann et al. (2014) for problem complexity under a variety of restrictions. There are also straightforward methods to build standard form ambiguity sets directly from sampled data (Van Parys et al., 2017).

2.4.4 Alternative Optimization Models

Alternatives to the standard stochastic and robust optimization models include fuzzy optimization and optimization under imprecise probabilities. Zimmermann

(1978) presents a comprehensive development of fuzzy linear programming. In a similar vein, Julien (1994) presents an approach to possibilistic linear programming. More recently, Doria (2017) provides foundational results for the properties of conditional previsions defined by Choquet integrals. The adoption of these non-probabilistic techniques in practice has been sparse, perhaps due to the comparatively small set of solvers and modeling tools as compared to the traditional optimization approaches.

2.4.5 Sequential and Adversarial Optimization

Dynamic decision theory has received significant attention in the recent literature (Klibanoff et al., 2009; Maccheroni et al., 2006; Siniscalchi, 2011). Often leveraging these results, optimization has also addressed sequential problems under uncertainty.

For short, finite horizons, two-stage and multi-stage problems allow the decision maker to take recourse actions after the uncertain parameters become known. These adjustable variables are functions of uncertain variables and these problems are commonly referred to as *adaptive optimization* problems (e.g., Bertsimas and Goyal (2010)). Two-stage and short-horizon multi-stage decisions can be modeled as a multi-level optimization problem. Sariddichainunta and Inuiguchi (2017) extend such a bilevel linear programming approach to the case with an uncertain lower level, using a convex polytope as the uncertainty set. Probabilistic decision graphs provide an alternative approach to sequential decision making in these settings (Jensen and Nielsen, 2013).

Arbitrary length and infinite horizon decisions are often modeled with Markov decision processes (MDPs) (Puterman, 2014). The class of uncertain MDPs includes a variety of approaches. Uncertain transition probabilities were introduced by Satia and Lave (1973). White and Eldeib (1986) and White and Eldeib (1994) also produced early uncertain MDP models, with similar work by Harmanec (2002) in the planning

literature. From the general class of MDPs with imprecise parameters (Delgado et al., 2011a, 2016), bounded parameter MDPs (Givan et al., 2000) and MDPs with set-valued transitions (Trevizan et al., 2007) were later developed as special cases in the artificial intelligence community. In the operations research literature, robust MDPs allow the parameters for the transition matrix to vary within uncertainty sets and use a worst-case solution approach (Wiesemann et al., 2013; Nilim and El Ghaoui, 2005; Iyengar, 2005). The high cost of robustness in this formulation motivates the distributionally robust MDP (Xu and Mannor, 2012), which replaces the uncertainty sets with ambiguity sets, as in the optimization literature, including standard form ambiguity sets specifically (Yu and Xu, 2016). Coherent risk measures have also been extended to the sequential decision making context in a similar manner (Artzner et al., 2007).

In addition to the uncertainty in transitions and rewards, in many real-world scenarios the environment is not completely observable. This leads to the need for partially observable MDPs (POMDPs). Although POMDPs can be reformulated as MDPs, the reformulation is non-trivial, which motivates the need for uncertain POMDPs in addition to uncertain MDPs. Early work developed the underlying theory and POMDP formulation (Dynkin, 1965; Aoki, 1965; Astrom, 1965) and these models have seen many applications in the following years (Monahan, 1982; Lovejoy, 1991; Aberdeen, 2003). The effort to incorporate uncertainty in MDPs has also been recently extended to POMDPs in the form of POMDPs with imprecise parameters (Itoh and Nakamura, 2007), robust POMDPs (Osogami, 2015), and ambiguous POMDPs (Boloori and Cook, 2017; Saghafian, 2018).

The most recent ambiguous POMDP model (Saghafian, 2018) addresses uncertainty in two unique ways. First, it combines model ambiguity with partial observability which allows for models that address problem settings with high levels of

uncertainty. Second, the key results are developed for both maxmin expected utility and α maxmin utility paradigms. Most other work in the optimization community focuses on the maxmin criterion. Further development of optimization results incorporating α maxmin, smooth ambiguity, or generalized uncertainty averse preferences would complement the existing body of work that focuses largely on expected utility and maxmin expected utility. Higher-order uncertainty has also been applied to adversarial models. The distinction between uncertainty sets and ambiguity sets is common in the game theory literature, resulting in robust stochastic games (Aghassi and Bertsimas, 2006) and distributionally robust stochastic games (Qu et al., 2017). In particular, Liu et al. (2018) use standard form ambiguity sets to take advantage of tractability results by Wiesemann et al. (2014). The connection to risk measures is also exploited by Loizou (2016). Data-driven ambiguity sets, chance-constraints, and moment-based ambiguity sets are also considered in the distributionally robust stochastic game literature (Sun and Xu, 2016; Ahipaşaoğlu et al., 2015; Singh et al., 2017). In static games, Sasaki (2017) considers the problem of stable generalized Nash equilibria in the presence of unawareness. These results demonstrate the utility of extending the foundational results in robust and stochastic optimization to sequential and adversarial problems.

2.5 Applications

In this section, we provide a brief overview of application areas for the major uncertainty models. Then, we present a more detailed review of applications of uncertainty sets and ambiguity sets, focusing on recent operations research literature from 2010 to early 2018.

2.5.1 Uncertainty Models

Typical applications of each major class of uncertainty models are shown in Table 2. Traditional Kolmogorov probability is used as the default probability model in most theory and practice, including the physical sciences, social sciences, economics, and decision making. However, finitely-additive de Finetti probability has also found some niche uses. For instance, it is used in some subjective decision making problems to ground the interpretation of the axioms in actual behavior (de Finetti, 1974). It has also been used to address the first digit problem, which deals with the phenomenon of the first digit of a given number in many application areas (e.g., insurance claims) being more likely to be low than high (Jech, 1992).

Table 2. Applications, adapted from Walley (2000)

Model	Typical Applications
Probability	Statistics, models of variability and error
Non-additive Probability	Subjective decision making, models of knowledge
Possibility Theory	Vague judgments of uncertainty in natural language
Evidence Theory	Multivalued mappings and non-specific information
Fuzzy Measure Theory	Statistical neighborhoods, economics
Imprecise Probability	Buying and selling prices for gambles, envelopes of expert opinions, preference judgments in decision making

Possibility theory, Dempster-Shafer theory, and fuzzy measure theory have all seen relative popularity in computer science applications. Possibility theory has been applied to scheduling, database querying, medical diagnosis, and risk analysis (Dubois and Prade, 2003). Dempster-Shafer theory has been applied to discriminant analysis in classification problems, information retrieval in database systems, and fusing data from different sensors (Smets, 1999). Fuzzy measures have been applied primarily to financial risk measures and risk analysis (Nguyen and Sriboonchitta, 2010). In addition to their intermediate use to develop some of the uncertainty models, fuzzy sets have been used directly for industrial control, industrial management, economic decision making, and inference in linguistic reasoning (Maiers and Sherif, 1985).

The three main types of imprecise probability are less widely used, but there are some examples of applications. They have been applied to economic decision making for solar energy, climate change, forestry, bio-medical problems, statistical learning, and reliability analysis (Miranda, 2008).

The broader approaches related to utility and sets have widespread application. VNM expected utility is prevalent across decision making applications, similar to the prevalence of traditional probability theory. Although not as ubiquitous as expected utility, prospect theory has seen broad application in finance, insurance, industrial organization, labor supply, and general decision making under uncertainty (Barberis, 2013; Wang and Guo, 2017; Siniscalchi, 2009; Arad and Gayer, 2012; Dimmock et al., 2015b; Li et al., 2017).

2.5.2 Optimization

Gabrel et al. (2014) review applications of robust and distributionally robust optimization. The applications include inventory and logistics (combinatorial problems, scheduling, facility location, inventory management), finance (portfolio optimization, risk measures, derivatives), revenue management (quantity management, quantity allocation, project selection, price management), stochastics (queuing networks, Markov decision processes, stochastic games), machine learning and statistics (maximum likelihood, support vector machines, principal component analysis), energy systems (oil supply chain, power grid, unit commitment), and the public good (medical treatment, patient management, humanitarian supply chain) (Gabrel et al., 2014).

Robust optimization continues to be a popular modeling technique in all of these areas of application. We highlight several representative applications, but these examples are not exhaustive. Recent examples of robust optimization include a surgery

scheduling model in the health care area (Neyshabouri and Berg, 2017). Logistics remains a popular application area with facility location, capacity acquisition, and technology choice (Jakubovskis, 2017), fleet sizing and routing (Lei et al., 2016), and supply chain planning models (Wong et al., 2016). Some recent approaches use dynamic programming to solve static robust optimization problems, for instance in robust lot-sizing (Agra et al., 2016; Santos et al., 2018) and robust knapsacks (Claßen et al., 2015).

Distributionally robust applications are becoming more prevalent, although still less widely used than robust optimization. Energy applications are disproportionately represented in the distributionally robust field, for example power generation unit commitment (Bai and Yang, 2014) and wind farm allocation (Alismail et al., 2018). There are also results in diverse areas such as network optimization (Nakao et al., 2017) and finance (Yang et al., 2014; Paç and Pnar, 2018; Ben-Tal et al., 2013). Yang et al. (2014) consider coherent risk measures, specifically conditional value at risk, and use box constraint uncertainty sets for an application to portfolio selection. Romanko and Mausser (2016) also consider value at risk optimization, but from a scenario-based perspective. Data-driven methods have also been studied recently in both robust and distributionally robust optimization (Bertsimas et al., 2018a; Mohajerin Esfahani and Kuhn, 2017; Gotoh et al., 2017; Delage and Ye, 2010; Wang et al., 2018; Gotoh and Uryasev, 2017). As in the case of robust optimization, this is set of references is representative but not exhaustive.

2.5.2.1 Robust Sequential Optimization

There are relatively few recent applications focusing on sequential decision making under ambiguity compared to those using unqualified Markov decision processes or robust optimization in a static setting.

Since the uncertain parameters in an MDP are often the transition probabilities themselves, the distinction between uncertainty sets and ambiguity sets is less clear in the sequential context. Typically, robust MDPs employ uncertainty sets, including for transition probabilities. Distributionally robust MDPs, however, typically make use of nested sets to build the ambiguity sets, especially the standard form ambiguity sets due to Wiesemann et al. (2014).

There are a variety of recent results that model uncertainty in robust sequential decision making. Wiesemann et al. (2013) formalize the framework for robust MDPs and the notion of (s, a)-rectangular uncertainty sets, which are referenced in several other applications. For instance, Gutin et al. (2015) present an application for interdicting project management schedules (terrorist plots, nuclear proliferation) using the (s, a)-rectangular sets. Coupled uncertainty sets generalize the uncoupled (s, a) rectangularity to k-rectangularity (Mannor et al., 2016). Ahmed et al. (2017) use rectangular uncertainty sets for both the reward and transition functions with a sampling approach for minimizing regret, with an application to stochastic inventory control.

The robust MDP applications are diverse, including data center controls (Weng et al., 2017), Kalman filters with interval uncertainty sets (Shashua and Mannor, 2017), and safety-critical systems (Dimitrova et al., 2016). Robust MDPs have also recently been used to address classic operations research problems, including asset pricing and the multi-item newsvendor problem (Ben-Tal et al., 2013), inventory control (Shapiro, 2011), lot-sizing (Tan and Hartman, 2011), and spectrum allocation (Wang et al., 2015b). In particular, Ben-Tal et al. (2013) use uncertainty regions defined by ϕ-divergences (e.g., chi-squared, Hellinger, Kullback-Leibler). Classic MDP solutions methods have also been extended to robust MDPs by Sinha and Ghate (2016) who use policy iteration for robust, non-stationary MDPs with arbitrary, nonempty,

compact uncertainty sets and rectangularity.

The standard robust MDP has also been extended with alternate modeling or solution techniques. Iancu and Trichakis (2014) account for Pareto efficiency within robust MDP solutions using polyhedral uncertainty sets with applications to portfolio optimization, inventory management, and project management. Similarly, Hahn et al. (2017) analyze multi-objective processes with interval uncertainty sets. Sinha et al. (2016) also use interval uncertainty sets, but their application is to response-guided dosing in the medical setting. Dimitrov et al. (2014) consider robust decomposable MDPs that are subject to a shared resource with applications to allocating school budgets. Influences from the computer science literature can be seen in robust POMDPs (Osogami, 2015), hidden parameter MDPs which use unusual uncertainty sets (Killian et al., 2017), and robust MDPs with learning (Lim et al., 2016). Computational approaches also influence the neural net approximate solution by Li and Si (2010) and the minimax regret solution by Oh and Kim (2011).

The control theory literature also employs robust MDPs. Bertuccelli et al. (2009) address robust MDPs from a controls perspective, mixing off-line robust planning with on-line adaptive planning. Ure et al. (2012) take a similar adaptive planning approach that updates the uncertainty sets through data collection. Other recent control applications include path planning (Wolff et al., 2012) and power systems (Xiao et al., 2012).

2.5.2.2 Distributionally Robust Sequential Optimization

There are also recent results that model uncertainty in distributionally robust sequential decision making. Yu and Xu (2016) present a standard distributionally robust MDP formulation that explicitly uses standard form ambiguity sets. Other ambiguity sets are also used, including conic confidence sets (Yang, 2017a) and norm-

based ambiguity sets with the Wasserstein distance (Yang, 2017b). As in the robust MDP case, rectangularity remains an area of attention (Shapiro, 2016), with further connections to coherent risk measures.

Classic operations research applications are also addressed in the recent distributionally robust literature. Bertsimas et al. (2018b) present adaptive distributionally robust MDPs using standard form ambiguity sets with applications to medical appointment scheduling and multi-period inventory control. Xin and Goldberg (2015) also consider distributionally robust inventory control, specifically when demand is a martingale, while Chen et al. (2018) consider data-driven inventory control with norm-based Wasserstein ambiguity sets. Energy systems continue to be of interest, with more applications to wind power (Samuelson and Yang, 2017) and hydrothermal scheduling (Huang et al., 2017).

There are several recent results that model uncertainty in POMDPs and MDPs using bounded-parameter methods (Ni and Liu, 2013; Fussuma et al., 2014; Scheftelowitsch et al., 2017), which correspond to interval uncertainty sets in the robust MDP literature. For the broader approach of MDPs with imprecise parameters, recent publications combine polyhedral uncertainty sets with a factored representation (Delgado et al., 2016, 2011b).

There has been limited overlap between the robust sequential decision making literature and behavioral decision theory. Liu et al. (2015) present an application to multi-period portfolio optimization with behavioral factors incorporated through prospect theory. They use a particle swarm optimization approximate solution technique to analyze market return data. Wang et al. (2015a) also use prospect theory to model a routing problem with interval uncertainty sets and a heuristic solution.

2.5.2.3 Adversarial Optimization

Robust and distributionally robust stochastic games have been applied to classic operations research problems with adversarial extensions, including the newsvendor problem and queues (Jiang et al., 2011; Kardeş et al., 2011). Lim et al. (2016) consider a partially adversarial problem where some uncertain parameters are adversarial and some are stochastic, using arbitrary compact uncertainty sets. Adversarial optimization under uncertainty has also been applied to security games that address terrorism (Kardeş, 2005, 2014), fisheries (Haskell et al., 2014), and other fields (Nguyen et al., 2014, 2016). The robust approach in optimization also has analogues in the game-theoretic research on *safe* strategies in competitive games (Johanson and Bowling, 2009; Ganzfried and Sandholm, 2015; Ponsen et al., 2011; Wang et al., 2011; Johanson, 2016).

2.6 Conclusion

This survey reviews recent approaches for addressing uncertainty through risk and ambiguity models. We develop a framework that describes the relationships among uncertainty modeling approaches for representation, uncertainty modeling approaches in decision making, and uncertainty modeling approaches in optimization. We also survey recent applications in optimization. These discussions provide a comprehensive view of uncertainty modeling from foundation to application and clarify the trade-offs among the alternative representations.

Uncertainty sets and ambiguity sets are the most formally developed representations of uncertainty in optimization and sequential decision making. Robust methods using uncertainty sets remain popular while distributionally robust methods using ambiguity sets have seen increasing application. Standard form ambiguity sets have been partially adopted but simpler representations also remain in use. Non-probabilistic

uncertainty models continue to see application outside of formal decision making, but have not been adopted in the optimization and sequential decision making communities.

The adoption of uncertainty modeling in sequential decision problems lags static optimization, despite the existence of tractable theoretical solutions. This suggests an opportunity for the development of improved techniques for sequential decision making under second-order uncertainty, which will likely increase adoption in practice and applied research. Recent advances toward generalizing ambiguity aversion in decision theory, such as smooth ambiguity and variational preferences, might be a useful foundation for extending sequential optimization under uncertainty.

There is also potential for further research at the intersection of behavioral decision making and robust optimization. Some recent results consider prospect theory decision models in a robust context, but advances in this area could lead to significant improvements in the applicability and realism of static and sequential decision models. As decision theory continues to advance and the experimental evidence base grows for attitudes toward ambiguity, there is opportunity to develop more sophisticated and practical optimization methods in static, dynamic, and multi-agent settings.

III. A Survey of Uncertainty Modeling in Operation Assessment

This chapter includes joint work with LTC Nicole Curtis, United States Army, who provided the descriptions of military assessment examples.

3.1 Introduction

Operation assessment is a critical aspect of modern military operations. Commanders, civilian leadership, and other stakeholders use assessment to determine progress toward goals at various levels of national security. Recent attention to operation assessment has generated discussion about how assessment practices can be improved across all types of security-related assessment.

Narrative and standards-based approaches are considered best practices in military operation assessment, but there is an opportunity to complement these approaches with best practices from the closely related field of evaluation theory (Arnhart and King, 2018). Some types of operation assessment are more directly related to evaluation theory than other types. There is a particularly close relationship for assessments dealing with development issues, such as assessment of peace operations, stability operations, counter-insurgency, or security cooperation. Assessment of operations with significant public opinion goals are closely related as well, including inform, influence, and persuade (IIP) operations (Paul et al., 2015b). Assessment of traditional kinetic and high-intensity operations differ significantly from standard program evaluations but they do have similarities with performance monitoring, a subset of evaluation theory.

While military assessment shares many characteristics with program evaluation and performance monitoring, the dynamic and extreme nature of conflict environments introduces some critical areas of dissimilarity. Not every assessment will be

able to make use of evaluation and monitoring approaches, and no assessment will be entirely programmatic in nature. Nonetheless, the theoretical foundations and practical guidelines developed over the past several decades of research and practice in evaluating and monitoring military operations in Afghanistan, Iraq, and elsewhere provide a rich source of material for assessors to use for improving the practice of operation assessment.

In the remainder of this chapter, we discuss the connections between evaluation theory and military assessment. In Section 3.2, we review a framework for classifying evaluations and discuss four major approaches for implementing evaluations. We also highlight different military contexts for each evaluation approach and propose a classification of military assessment along two dimensions of evaluation (formative-summative and qualitative-quantitative). In Section 3.3, we describe three classes of evaluation design (descriptive, quasi-experimental, and experimental) with corresponding military applications. In Section 3.4, we review qualitative and quantitative uncertainty representations in evaluation and assessment. We conclude by summarizing the relationship between evaluation theory and military assessment and by suggesting future areas of research.

3.2 Evaluation, Monitoring, and Military Assessment

3.2.1 Evaluation Theory

The primary purpose of evaluation is to determine the value or quality of an organization, program, or policy (Fitzpatrick et al., 2011). This primary purpose directly enables related purposes including supporting decision making, program improvement, accountability, and long-term organizational improvement (Fitzpatrick et al., 2011). All of these secondary purposes are relevant in the national security context, but military assessments often emphasize decision making and accountability.

For each of these purposes, there are also different approaches for conducting evaluations. Wholey et al. (2010) suggest several continuous dimensions which define the range of possible evaluation types, as shown in Figure 4. Fitzpatrick et al. (2011) organize the different approaches to evaluation by grouping them into four major categories: expertise-oriented, program-oriented, decision-oriented, and participant-oriented. These categories represent alternative ways of framing the context and purpose of evaluation. Assessors may find different categories appropriate for different assessment problems or find that a blended approach is most appropriate. We consider each of the four approaches in more detail in the following subsections.

Figure 4. Dimensions of evaluation, adapted from Wholey et al. (2010)

3.2.1.1 Expertise-oriented Evaluation

Expertise-oriented evaluation is the earliest and most straightforward type of evaluation (Fitzpatrick et al., 2011). In this evaluation paradigm, an expert or group of experts studies a program or organization and judges its quality.

Fitzpatrick et al. (2011) further classify expertise-oriented approaches as formal review systems, informal review systems, ad hoc panel reviews, and ad hoc individual reviews. Formal review systems have a structured review process, published standards, multiple expert reviewers, and the ability to affect the status of the organization under review (Fitzpatrick et al., 2011). At the other end of the spectrum, ad hoc individual reviews lack both formal structure and published standards and

are often conducted by a single reviewer without significant ability to change the organization or process under evaluation.

Although the formal review system is the most comprehensive of the expertise-oriented approaches, each of the less developed approaches can be appropriate under certain circumstances. For instance, a single expert brought in as a consultant can provide meaningful and expedient assessment of basic processes. Whether ad hoc or formalized, expertise-oriented evaluation relies completely on highly competent and objective reviewers. The reviewer plays the role of a critic who describes, interprets, and evaluates the system under observation (Eisner, 1976). The primary contribution of the critic is not the final judgment, but the ability to highlight important aspects of the program or system which may not be clear to non-expert observers (Fitzpatrick et al., 2011).

In the development of new weapon system capabilities, military analysts have used wargaming as a method to develop operational concepts, tactics, techniques, and procedures. These wargames often use vignettes within an approved Defense Planning Scenario to provide qualitative assessments and observations that are later integrated into either a closed-form or human-in-the-loop simulation (N. Curtis, personal communication, July 10, 2018). Another example of expertise-oriented evaluation is a Rehearsal of Concept (ROC) Drill which assessors use to examine the flow of forces into an area of operations during a crisis (N. Curtis, personal communication, July 10, 2018). Both methods rely on several subject matter experts across multiple functional roles to help commanders refine requirements and concepts while identifying risks.

By relying on expert judgment, the expertise-oriented approach addresses holistic factors that may be difficult to consider with other approaches. However, the method is also vulnerable to personal biases and lack of evaluation expertise. While orga-

nizations often use this approach in practice and academia has formally studied it, modern evaluation theory typically does not recommend expertise-oriented evaluation (Fitzpatrick et al., 2011).

3.2.1.2 Program-oriented Evaluation

Program-oriented evaluation shifts the focus from expert reviewers to objective standards based on the goals of the program. Program-oriented evaluation includes both objective-oriented evaluation and theory-driven evaluation as sub-categories. In the objective-oriented approach, evaluators specify explicit program goals, identify measures which indicate whether those goals are met, then collect and analyze data related to the measures (Tyler, 1942). The U.S. government and the Department of Defense have conducted objective-oriented evaluation in a variety of contexts (Fitzpatrick et al., 2011; National Performance Review, 1993; Office of Management and Budget, 2004). This classic approach to evaluation remains popular but has been "largely discredited by professional evaluators today" (Fitzpatrick et al., 2011, p. 155). The major limitation of objective-oriented evaluation is that it does not enable the evaluator to determine why a program is succeeding or failing (Fitzpatrick et al., 2011). A secondary limitation is that only the identified objectives will be considered in the final judgment, which will necessarily be incomplete (Fitzpatrick et al., 2011).

Theory-driven evaluation, also called theory based evaluation, addresses this concern. In theory driven evaluation, the evaluator explicitly records the logical structure of the program. There are various methods of defining the program theory, but a common approach is to define a logic model that includes inputs, actions, outputs, and outcomes (Fitzpatrick et al., 2011). There is a critical distinction between outputs, which are the immediate products of the program, and outcomes, which are the de-

sired results, categorized as near, medium, or long term (Fitzpatrick et al., 2011). Theory-driven evaluation also requires measurement of the implementation environment (Fitzpatrick et al., 2011). By measuring both outcomes and the environment, evaluators can determine whether the program was implemented as designed. If not, it represents a failure of implementation. If it was implemented according to plan but the desired results are not achieved, the program theory may be inadequate or the environment may be different than expected. This approach helps the evaluator to identify why the program is succeeding or failing.

Objective-oriented evaluation is a common approach in military assessments, exemplified by standards-based assessment (Arnhart and King, 2018). Assessors often augment the standardized bins in this approach with a written narrative that captures additional nuance and risk to account for the limitations of objective-oriented methods. Recent military assessments have also used theory-driven evaluation. Marquis et al. (2016) develop an assessment framework for defense security cooperation using an evaluation theory foundation. Paul et al. (2015a) also use a theory-driven framework for IIP evaluation. In official guidance, Department of Defense instructions explicitly recommend the use of evaluation theoretic concepts in defense assessments (Office of the Under Secretary of Defense for Policy, 2017). Conversely, the evaluation community has also recognized the link between military assessment and evaluation (Williams and Morris, 2009).

3.2.1.3 Decision-oriented Evaluation

While similar to program-oriented evaluation, the decision-oriented evaluation approach focuses on producing evaluations that impact decisions. A rigorous evaluation is not useful if its recommendations are not implemented, but there are many challenges associated with translating an evaluation report into action. These challenges

include methodological factors and the desirability of the findings, but the two most important factors are political considerations and the presence of a senior leader who supports the evaluation (Patton, 1994). Given these findings, utilization-focused evaluation primarily focuses on identifying and integrating a senior decision maker into the evaluation process (Patton, 2008). Although this support can make the evaluation more likely to successfully influence decisions, it can be difficult to find a senior leader with both the interest and the power to support the evaluation (Patton, 2008). Furthermore, organizational decision making typically involves multiple decision makers and bureaucratic systems, which may limit the positive impact of a single decision maker involved in the assessment process (Weiss and Mark, 2006).

Although theory-driven evaluation has largely replaced objective-oriented evaluation, performance monitoring is still relevant today as a decision-focused evolution of classic objective-oriented evaluation. Performance monitoring continuously collects data on outputs and outcomes with a focus on program improvement (formative evaluation) (Fitzpatrick et al., 2011). This type of monitoring is limited to data that is easy and inexpensive to collect, so it often focuses on a small set of quantitative measures (Fitzpatrick et al., 2011). However, performance monitoring plays a useful, complementary role to traditional theory-driven evaluations, which typical provide feedback at a slower rate.

Performance monitoring is particularly useful in a military context at the tactical levels or for time-sensitive decision making. Ideally, analysts can automate this type of data collection for presentation in a dashboard or other continually accessible format. The data reporting mechanism should include a quantitative estimate of uncertainty along with the reported value. There is also overlap between performance monitoring from an assessment perspective and routine military intelligence gathering.

Military commanders, who are responsible for every Soldier and civilian assigned

to the installation, often use decision-oriented evaluations when assessing installation "campaign plans" (N. Curtis, personal communication, July 10, 2018). Similar to campaign plans for military operations, these installation campaign plans have specific objectives and tasks with associated metrics that assess outcomes with respect to a defined end state (N. Curtis, personal communication, July 10, 2018). This continuous monitoring allows commanders to shift resources to activities that need improvement. Installation level campaign plans include programs such as Soldier Sponsorship and Inprocessing, Partners in Education, and Suicide Prevention (N. Curtis, personal communication, July 10, 2018).

Another example of the military's use of decision-oriented evaluation is in suicide prevention training effectiveness assessment. One Army installation funded a civilian suicide intervention training program as part of their suicide prevention efforts (N. Curtis, personal communication, July 10, 2018). Upon completion of the two-day workshop, participants completed a survey which primarily provided results to the contractor who developed the program (N. Curtis, personal communication, July 10, 2018). The installation commander directed an assessment of the training in which analysts examined over 11,000 participants' data (N. Curtis, personal communication, July 10, 2018). The positive feedback and high level of participant satisfaction resulted in recommendations for continued program funding (N. Curtis, personal communication, July 10, 2018). Additionally, the results helped the program vendor tailor the instruction for a military audience and informed the installation's higher headquarters of an effective alternative to the current mandated program (N. Curtis, personal communication, July 10, 2018).

3.2.1.4 Participant-oriented Evaluation

Participant-oriented evaluation differs significantly from program-oriented and decision-oriented evaluation. There are many variants of participant-oriented evaluation but, in all of these approaches, participants have power to influence the evaluation. Practical participatory evaluation (Cousins and Earl, 1992) and stakeholder-based evaluation (Mark and Shotland, 1985) are two of the most relevant approaches to defense assessments. Fitzpatrick et al. (2011) characterize participatory evaluation using three primary criteria: the degree to which participants have control over the evaluation process, the breadth of stakeholders involved in the process, and the depth of participation of those stakeholders. Practical participatory evaluation and stakeholder-based evaluation both moderately expand the control, breadth and depth of stakeholders while maintaining significant control with the evaluator (Fitzpatrick et al., 2011). Stakeholders assist in determining the overall structure of the evaluation and in collecting and analyzing data. The evaluator's role includes acting as a technical consultant and negotiator among the stakeholders (Fitzpatrick et al., 2011).

There are several strengths and limitations for participant-oriented approaches. They involve the end users of the program and may make eventual adoption of the evaluation recommendations more likely (Fitzpatrick et al., 2011). This approach is one way to address the challenge of political or bureaucratic factors limiting an evaluation's effectiveness. Sustained participant involvement can also enhance the evaluator's understanding of the program, leading to a better evaluation (Fitzpatrick et al., 2011). However, participant-oriented approaches can be difficult to implement if there are many stakeholders or if the stakeholders are difficult to access. Senior decision makers may also perceive participatory evaluation as illegitimate due to bias or incompetence (Fitzpatrick et al., 2011). The participants who plan and execute the program are involved in the evaluation, which may bias the results. The participants

are also not trained in evaluation, so their increased control over the evaluation process may degrade the rigor of the results.

Military analysts have used participant-oriented evaluation to assess brigade-level home station training. On one Army installation, analysts assessed a new capability that provided standardized home station training through live, instrumented, and collective training experiences (N. Curtis, personal communication, July 10, 2018). Stakeholders at the Army Service Component Command participated in developing the scope of the analysis and the Soldiers conducting the training provided feedback through questionnaires, focus group sessions, and key leader interviews (N. Curtis, personal communication, July 10, 2018). Participants' perceptions of the training experience informed the analysis that provided recommendations for continued funding and implementation of an instrumented home station training capability (N. Curtis, personal communication, July 10, 2018).

3.2.2 Military Assessment

Military assessments take a wide variety of forms which can be categorized by each of the four approaches discussed previously and along the six dimensions of evaluation. Although practice often deviates from established doctrine and policy, official military guidance is available through Department of Defense instructions and military doctrine at the service, joint, and coalition levels (Office of the Under Secretary of Defense for Policy, 2017; North Atlantic Treaty Organization, 2015; United States Joint Chiefs of Staff, 2017a,b; Joint Staff J-7, 2011; United States Joint Chiefs of Staff, 2015b,a; Office of the Chief of Naval Operations, 2013; Department of the Army, 2014, 2017). A partial list of military assessments includes:

- national strategic assessment
- country assessment

- security cooperation program assessment
- global campaign assessment
- contingency campaign assessment
- theater campaign assessment
- operational campaign assessment
- operation assessment
- intelligence assessment
- training assessment
- operating environment assessment
- after-action review (Military Operations Research Society, 2018).

Using the dimensions of evaluation developed by Wholey et al. (2010), the majority of military assessments are problem oriented and goal-based. Conversely, most military assessments can be either ongoing or one-shot and either objective or participatory. However, the remaining two dimensions provide a useful distinction among the different types of assessment applications. *Formative evaluation* focuses on improving the task or operation which is under assessment, whereas *summative evaluation* focuses on evaluating the effectiveness of a task or operation after it is complete (Wholey et al., 2010). *Quantitative* methods analyze numerical data, while *qualitative* methods address narrative responses, images, and other unstructured data (Wholey et al., 2010). As Figure 5 indicates, assessments can employ blended or intermediate methods along each dimension. The distinction between quantitative and qualitative methods, in particular, has diminished as modern analytic methods become more sophisticated at addressing unstructured data. For instance, text mining of open-ended

survey responses can produce a quantitative sentiment analysis of unstructured narrative data. The distinction between quantitative and qualitative data and methods is not always clear, and most assessments use both types of data and both types of methods.

Figure 5 shows a notional, relative categorization of each of the military applications along the two primary dimensions. Note that the precise location of each assessment type is not fixed and may change depending on context, but the overall relative relationships suggest different areas of emphasis for each type of assessment.

Figure 5. Types of military assessments

3.3 Assessment Design and Analysis

Within any of the overarching evaluation approaches discussed previously, selecting the assessment design is a key part of planning an effective assessment. The major classes of assessment designs are descriptive, quasi-experimental, and experimental (Fitzpatrick et al., 2011). The design choice directly affects the conclusions that can

be drawn from the assessment. The underlying trade-off is that designs supporting causal claims are usually more expensive and logistically difficult to implement. Assessment resources may be limited with respect to time, available personnel, or funding. The resource constraints influence the type of assessment than can be implemented. If there are severe restrictions on the time available for evaluation, a hasty assessment might use only qualitative, narrative methods. If there are few restrictions on the time available for evaluation, a deliberate assessment can build a full evaluation of the operation using robust quantitative and qualitative methods. If a full experimental design is not feasible, but descriptive or quasi-experimental evidence supports a conclusion, Wholey et al. (2010) suggest using the term *plausible attribution* instead of stronger causal terms. The following subsections discuss each major class of assessment design in more detail, drawing connections between Chapter 15 (Fitzpatrick et al., 2011) and military assessment.

3.3.1 Descriptive Designs

Descriptive designs are a comparatively low cost way to collect important information on the state of the program and its outputs and outcomes. While descriptive designs cannot answer strategic questions related to root causes, they are critical for providing a basic level of understanding for the assessment. If a descriptive design indicates a desired outcome is not being achieved, it may avoid the need for a lengthy causal assessment. Descriptive designs are particularly useful for performance monitoring. The main types of descriptive designs are case studies, cross-sectional designs, and time-series designs.

3.3.1.1 Case Study

Practitioners frequently use case study methods. These designs allow for an in-depth exploration of a single or small number of cases but do not provide enough information to generalize to other environments (Fitzpatrick et al., 2011). Case studies can use both quantitative and qualitative methods but typically focus on qualitative methods. The primary qualitative methods are observations, interviews, and the study of existing documents (Fitzpatrick et al., 2011). In some cases, quantitative surveys or statistical analysis of existing data is also appropriate (Fitzpatrick et al., 2011). The output of a case study design is typically a narrative report providing a detailed discussion of the case or cases studied. Assessors select the cases under study because they are either typical or exceptional, whether successful or unsuccessful (Fitzpatrick et al., 2011). If the operation under assessment is small enough, a case study may provide all the needed information. Stake (1995) and Yin (2009) provide detailed guides for implementing case studies.

While case studies are primarily descriptive, Yin (2009) argues that case studies can play a critical role in causal investigations. Experiments often impose artificial restrictions on the number of variables investigated or the environment in which the study is performed (Yin, 2009). Case studies address these problems directly by investigating the full complexity of a real-world case. Case studies are rarely sufficient for making causal conclusions, but they can play an important complementary role (Cook et al., 2002).

In the military context, assessors frequently use a case study approach. A case study may select one geographic area to study in depth using phone interviews with operational forces and existing after action reports (AARs). The case study sometimes includes specific interview questions or quantitative survey questions on the AAR form. If possible, the assessor might travel to the location or unit under study

to get a better understanding of the performance of the unit.

Assessors can also use case studies preemptively with trial runs. For instance, the results of an ROC Drill that assessed the flow of military units into the theater of operation primarily provided qualitative feedback to the Combatant Commander (N. Curtis, personal communication, July 10, 2018). This feedback was similar to an AAR, highlighting specific issues or changes in the environment not known during the development of the deployment plan. Specifically, poor runway conditions in a partner nation highlighted a need for repairs that would cause a delay in large cargo aircraft arrival into theater (N. Curtis, personal communication, July 10, 2018). This insight led to a change in the sequence of equipment and personnel arriving in theater and planners then used the results of the ROC Drill to refine the plan and associated deployment models (N. Curtis, personal communication, July 10, 2018).

Another example of a case study in a military context is the use of vignettes, based on a strategic-level Concept of Operations or Defense Planning Scenario (DPS) (N. Curtis, personal communication, July 10, 2018). Because these are strategic-level operations, analysts examining new weapon system capabilities often develop vignettes appropriate to the scope of the assessment. An assessment of a new unmanned aerial system controlled by a company formation may only need to evaluate its capabilities within a battalion-level mission set (N. Curtis, personal communication, July 10, 2018). Focusing on tactical operations and specifically decisive action, "the continuous simultaneous combination of offensive, defensive, and stability or defense support of civil authorities' tasks" (Department of the Army, 2017) requires a greater level of detail than what is found in the DPS. Due to resource and time constraints, these constraints may limit the number of vignettes considered by the assessment.

3.3.1.2 Cross-Sectional

The cross-sectional design is a basic design used to collect information across a variety of subjects at a specific point in time (Fitzpatrick et al., 2011). Cross-sectional designs typically use surveys on opinions and behaviors, but they can also employ other data collection techniques (Fitzpatrick et al., 2011). The design can include all subjects of interest or a subset of the full population. The design typically breaks subjects into interesting sub-groups (Fitzpatrick et al., 2011). It is important to include enough questions or data items to identify relevant sub-groups. Given the broad nature of cross-sectional designs, they can identify problems or help to set priorities for further exploration in case studies (Fitzpatrick et al., 2011).

Military applications frequently use this type of design. Using a cross-sectional design is valuable when comparing training feedback from diverse populations. For example, the perceived quality of a military training event may vary between junior enlisted soldiers in an infantry company and the junior enlisted soldiers in an infantry brigade headquarters due to their markedly different operational roles (N. Curtis, personal communication, July 10, 2018). Likewise, a junior enlisted soldier will have a different experience than a senior officer (N. Curtis, personal communication, July 10, 2018). Differentiating among multiple military occupational specialties, ranks, and units is useful to understand variance in the data and allows analysts the opportunity to tailor recommendations for specific populations. Figure 6 displays this type of cross-sectional data by sub-group using survey data from the Iraq Index (O'Hanlon and Campbell, 2007). The figure communicates uncertainty in the category values with error bars representing the 95% confidence intervals for each sub-group, which show more uncertainty in the Kurdish polling data than in the data for the other population groups. Note that this continuous summary would not be appropriate for ordinal data. The major limitation of cross-sectional designs is that they do not

provide descriptive data for trends over time.

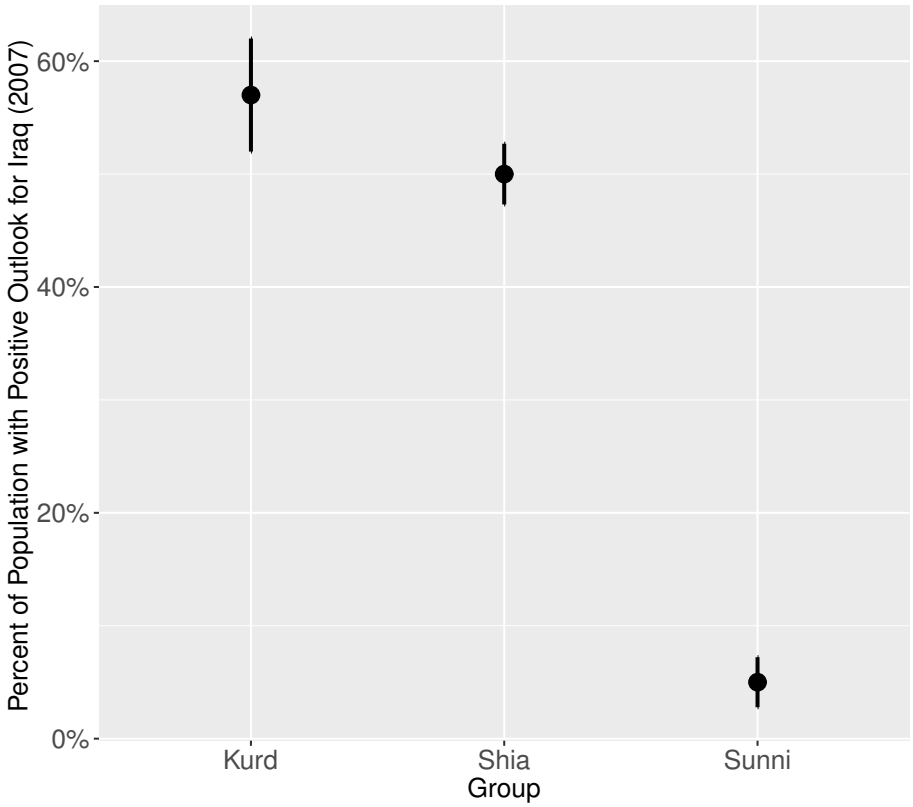

Figure 6. Cross-Sectional data with uncertainty (O'Hanlon and Campbell, 2007)

3.3.1.3 Time-Series

Time-series designs, also called longitudinal designs, provide the data necessary to describe trends over time (Fitzpatrick et al., 2011). The key decisions in a time-series design are the frequency of collection and the total amount of data to collect (Fitzpatrick et al., 2011). It is important to ensure that the data quality is consistent over the design time frame, so any trends identified reflect changes in the environment rather than changes in data collection methodology (Fitzpatrick et al., 2011).

While time-series designs are relatively straightforward, the interpretation of time-series data can be difficult, even without considering causation. In some cases it may be difficult to distinguish between noise and a trend in the data. Seasonality is also

common in time-series data and may be difficult to identify, particularly if there are multiple seasonal components with different cycles.

The major barriers to time-series designs in military operation assessment are the dynamic nature of operational objectives and the frequent turnover of personnel. Different decision makers may be interested in different strategic questions, or the mission may change rapidly enough that historic data are no longer relevant. If available, pre-existing historical data can mitigate these challenges. Despite these challenges, time-series designs are common and provide a useful descriptive component in most assessments. When presenting multiple descriptive time-series together, it is important to accurately represent the limitations for assigning causation. Figure 7 shows a sample presentation of time-series data. This plot shows smoothed data with uncertainty represented by a 95% confidence interval, using Iraq Index data for national electricity generation (O'Hanlon and Campbell, 2007). The uncertainty level for this data set is relatively low and stable. In some cases, the raw data may be more appropriate.

The need for baseline data in time-series analysis combined with the frequently shifting goals in a military context suggest that repeated data collection on indicators not currently of interest, but which may become important, is a useful task. Clearly, data collection and storage have associated costs, but analysts should consider both the current utility and the potential utility of data for future assessments in the cost-benefit analysis.

3.3.2 Quasi-Experimental Designs

Although describing the operational effects and the operating environment is a critical part of assessment, strategic questions often include a causal component. For instance, one of the strategic questions cited in Arnhart and King (2018), "To what

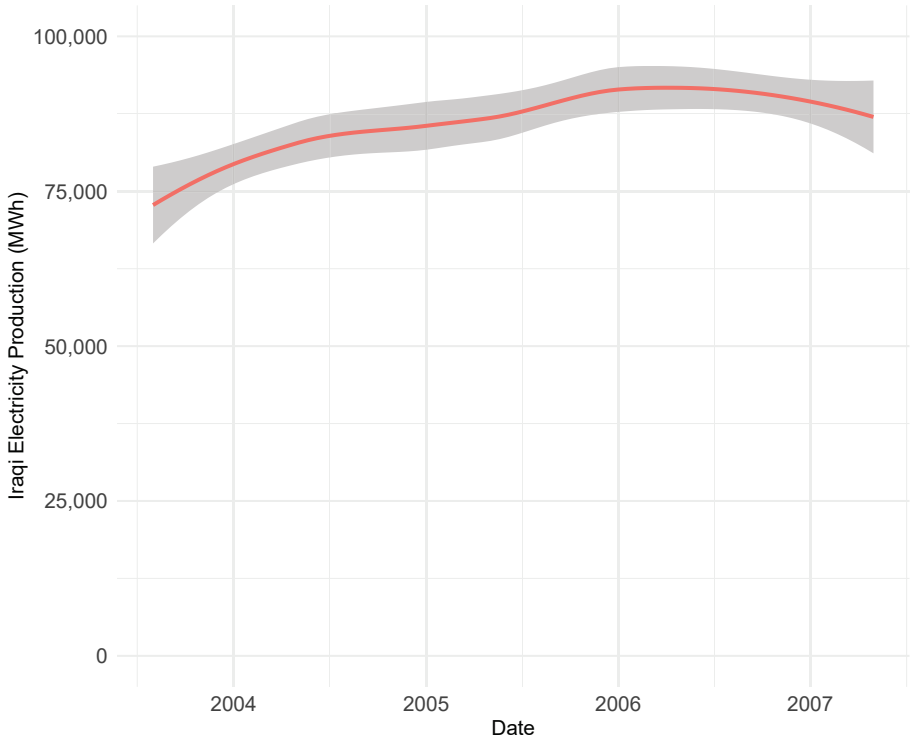

Figure 7. Time-Series data with uncertainty (O'Hanlon and Campbell, 2007)

extent have military operations deterred the actions of terrorist groups?", seeks to determine whether terrorist groups were deterred (descriptive) and if military actions caused the deterrence (causal). Unfortunately, causal designs are typically more difficult to implement. This difficulty motivates the use of designs that provide more information than descriptive designs without incurring the full costs of a causal design. Although the terminology of "descriptive" and "causal" suggests a binary distinction, the rigor and strength of assumptions associated with a design are continuous attributes. The main types of quasi-experimental designs are comparison group designs and interrupted time-series designs.

3.3.2.1 Comparison Group

The comparison group design is an extension of the cross-sectional design. This design compares subjects exposed to the operation or program under study with sub-

jects that have not been exposed (Fitzpatrick et al., 2011). However, the design selects quasi-control subjects from pre-existing groups (e.g., neighboring regions) rather than selecting them randomly. Ideally, the experimenter will measure both the target group and the quasi-control group on relevant indicators, both before and after the operation or program (Fitzpatrick et al., 2011). The target group and quasi-control group should be similar with respect to the critical pre-intervention measures. Collecting additional descriptive data about both groups can also help to establish their similarity (Fitzpatrick et al., 2011).

The major limitation to the comparison group method is that there may be systematic differences between the two groups. If the two groups have different outcomes after the intervention, the assessor will not be able to determine whether the difference is due to the intervention or to the underlying group differences (Fitzpatrick et al., 2011). The groups may have differences not captured in the indicators and descriptive measures, or they may start the same but diverge over time (Fitzpatrick et al., 2011). This type of design is also vulnerable to spill-over effects, where the primary operation incidentally affects the quasi-control group (Fitzpatrick et al., 2011).

In the dynamic military environment, it can be difficult to identify a suitable quasi-control group. Geographic regions provide a natural grouping if there is sufficient similarity between groups and there are few spill-over effects. Figure 8 provides a sample presentation of comparison group results, using survey data from an ABC and BBC news poll (ABC News, 2007). The presentation is similar to the cross-sectional presentation, but with each category split into the primary and comparison groups. In this analysis, there is more uncertainty in the Southwest Afghanistan region because it relies on a smaller subset of data than the combined regions. As in the cross-sectional case, this type of display is only appropriate for continuous data. There is a wide variety of other data visualization options for both continuous and

ordinal data (Wickham, 2016).

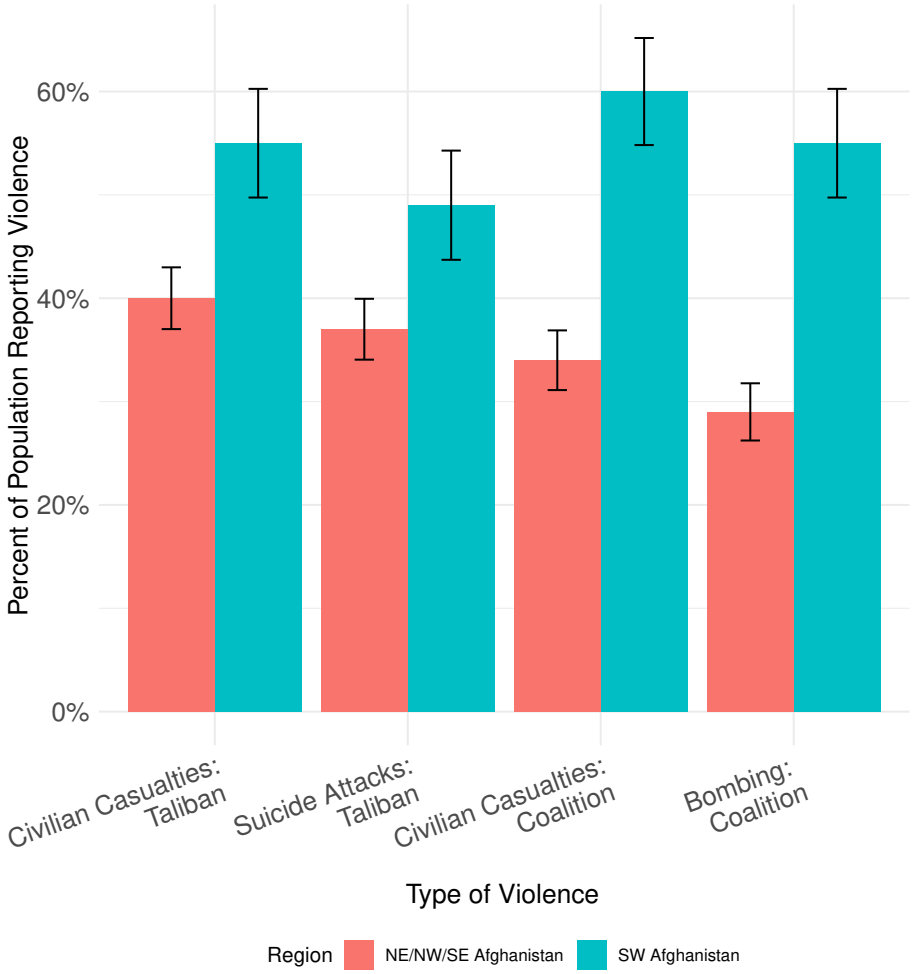

Figure 8. Comparison group data with uncertainty (ABC News, 2007)

3.3.2.2 Interrupted Time-Series

The interrupted time-series design requires repeated data collection before and after an event of interest (Fitzpatrick et al., 2011). This type of design often takes advantage of pre-existing historical data. It is important to have a sufficient length of time-series data collected prior to the intervention. Military plans often target operations to address problems which have recent poor performance, so if the historical time-frame is too short, the interrupted time-series design may falsely show

improvement when the data are varying within usual parameters (Fitzpatrick et al., 2011). If the characteristics of the time series data taken after the event are different than the characteristics of the time series data collected before the event, there is some evidence that the event impacted the data. Additionally, the time-series may differ in average value, trend, seasonality, or noise.

However, some other synchronous, outside factor may have also impacted the data. This concurrent factor may be an uncontrolled environmental factor or other synchronous operations. Another complicating factor in interpretation is that the event under study is often not instantaneous and the impacts of the event may be lagged (Fitzpatrick et al., 2011). Similarly, the impacts of other environmental factors may have lagged effects on the post-event time-series. These limitations make interrupted time-series a quasi-experimental design rather than a full experimental or causal design.

Interrupted time-series designs are a form of comparison with a baseline, where the pre-event time-series can be considered the baseline performance. For instance, to demonstrate value in a country assessment framework, the time-series value of a meaningful outcome indicator before and after a funding program provides some evidence that the funding led to improved outcomes (although still not definitive evidence). Reporting the value of the indicator only after the funding provides much weaker evidence of improvement due specifically to the funding. Figure 9 provides an example of the results from an interrupted time-series of civilian casualty data from Iraq (O'Hanlon and Campbell, 2007; Conflict Casualties Monitor, 2018). The dashed red line indicates the deployment of troops under the "surge" strategy, which coincides with a sharp decline in civilian casualties. Since this analysis is quasi-experimental, we cannot definitively attribute the decline in violence to the surge but the evidence suggests a plausible relationship (Saie and Ahner, 2018). We discuss the issue of

uncertain attribution in further detail in Section 3.4.1. Note that this plot shows raw data, but a smoothed version similar to Figure 7 is another display option.

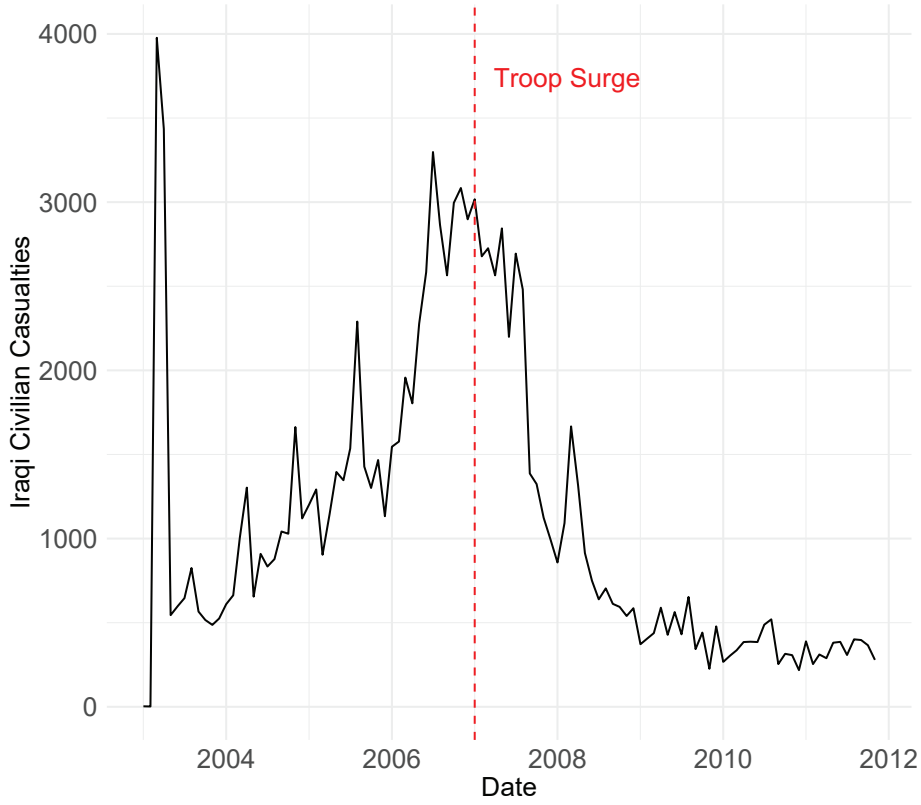

Figure 9. Interrupted time-series data (Conflict Casualties Monitor, 2018)

Operation assessments examining policy implementation and effectiveness may use interrupted time-series. Military commanders often track data derived from sexual assault and sexual harassment reports. Data elements include metrics such as the number of incidents, the type of report, number of incidents involving alcohol, gender of victim, and the time of day the incident occurred (N. Curtis, personal communication, July 10, 2018). In some instances, there may be several years of data. With historical data, analysts can assess the impacts of the new policy by examining data before and after the policy was enacted. Analysts must also be cautious in using historical data because the data collection process or the definition of data elements may have changed over time.

3.3.3 Experimental Designs

Experimental designs provide stronger evidence than quasi-experimental designs or descriptive designs. The purpose of an experiment is to determine the causal relationship between a suspected cause and effect (Cook et al., 2002). The ideal causal test is to compare reality to the counterfactual, an alternate reality without the suspected cause (Cook et al., 2002). Because a true counterfactual is impossible, experiments and quasi-experiments attempt to approximate the counterfactual. The primary distinction between experiments and quasi-experiments is that experiments use randomization (Cook et al., 2002). Randomization is a key feature of experiments because it protects against unknown biases. See Cook et al. (2002) for a detailed discussion of causal inference and the distinction between experiments and quasi-experiments. Many strategic questions seek to determine causes, which leads to experimental designs as the strongest form of evidence available.

In addition to the cost and logistical difficulty of experiments, sometimes experiments are inappropriate if assessors expect the operation or program to be much more effective than current operations. In this case, purposefully limiting the scope of the operation to gain more information about the causal question may be unattractive. There is a delicate balance between experimentation to determine what is effective and exploitation of methods known or expected to be effective. The primary risk of fully exploiting a promising alternative without experimentation is that expectations of effectiveness might turn out to be incorrect. The main types of experimental designs are posttest-only designs and pre-post designs.

3.3.3.1 Posttest-only

The posttest-only design randomly assigns subjects to treatment groups and compares them against a pre-determined measure after treatment (Fitzpatrick et al.,

2011). Some groups may include no treatment or the status quo treatment (Fitzpatrick et al., 2011). In the simplest case, there is a comparison between a treatment and no treatment, but some experiments may involve several different factors with varying treatment levels. In this case, analysts should use *design of experiments* to efficiently collect information on all factors simultaneously. For more details, see the work by Montgomery (2017).

Although referred to as a posttest, assessors can use surveys, interviews, observations, or data collection to derive the comparison measure and may also include multiple measures (Fitzpatrick et al., 2011). It is important to ensure treatment groups are kept as independent as possible during the operation or program, so the treatment for one group does not affect another group (Fitzpatrick et al., 2011).

Test and evaluation frequently uses this type of assessment, but it also relevant for military assessment. In the suicide intervention training program one Army installation implemented, Soldiers completed a 20-question survey at end of the workshop (N. Curtis, personal communication, July 10, 2018). Of the 11,000 soldiers who participated in the workshop, analysts compiled approximately 2,700 responses, randomly selected over a 14-month period (N. Curtis, personal communication, July 10, 2018). Soldiers answered questions requiring open-ended responses as well as questions that included Likert-style agreement scales (e.g., strongly disagree, disagree, neutral, agree, strongly agree) (N. Curtis, personal communication, July 10, 2018). These questions measured overall satisfaction of the training and perceived preparedness to intervene when someone expresses suicidal ideation (N. Curtis, personal communication, July 10, 2018).

3.3.3.2 Pre-Post

The pre-post experimental design is similar to the posttest-only design, but it adds a pretest for the treatment and control groups (Fitzpatrick et al., 2011). Again, the pretest terminology refers generically to any measure or set of measures taken before the operation or program. Although the design randomly assigns subjects to groups, the groups are sometimes small or lose members over the treatment period. The purpose of the pretest is to establish equivalence between the groups on the measures considered (Fitzpatrick et al., 2011). The pretest can also provide data for a statistical correction for changes in each group's composition over time (Fitzpatrick et al., 2011).

In the data analysis phase, this design supports a comparison between the posttest results for the treatment group and the control group. For operations or programs with lengthy implementation periods, it does not support a comparison between the pretest and posttest of the treatment group (Fitzpatrick et al., 2011). The change from the pretest to the posttest may be caused by the treatment or by outside factors. The same limitations that apply to interrupted time-series apply here, but magnified by the restriction to a single pretest measurement and a single posttest measurement (rather than a full time-series).

Although this is the most rigorous test design and provides the strongest evidence for answering causal strategic questions, the costs in terms of time and funding are high. In some cases, when the strategic question is critical to the mission and there are sufficient resources and support, a pre-post design may be appropriate. In some domains, for instance IIP operations, it may be more feasible to conduct such designs. Digital social media services are amenable to random selection of groups to receive different information campaigns and it can be relatively easy to collect opinion and demographic data before and after the operation.

3.3.4 Assessment Feasibility

The designs discussed in the previous sections are usually costly to implement. In some cases, the value the assessment provides to decision makers does not justify the organizational effort required to conduct the assessment. This problem is particularly relevant for summative assessments. Wholey et al. (2010) suggest that prior to beginning an assessment, the following criteria should be met:

- program goals are agreed on and realistic
- information needs are well defined
- evaluation data are obtainable
- intended users are willing and able to use the evaluation information.

If these criteria are not met, assessors can use exploratory evaluation to refine goals and information needs and to build support for a full assessment. Table 3 summarizes several approaches to exploratory assessment, with further details available in the work by Wholey et al. (2010).

Table 3. Exploratory evaluation approaches, adapted from Wholey et al. (2010)

Approach	Purpose	Staff Months
Evaluability assessment	Evaluate readiness, clarify goals, focus evaluation	0.5 to 3
Rapid feedback evaluation	Estimate effectiveness and uncertainty, produce tested designs, focus evaluation	3 to 12
Evaluation synthesis	Synthesize prior studies	1 to 3
Small-sample studies	Estimate effectiveness, produce tested indicators	0.5 to 12

3.4 Uncertainty in Assessments

Each of the designs discussed in Section 3.3 introduces uncertainty into the assessment in different ways. Measuring and communicating uncertainty is a difficult but critical part of operation assessment. In an engineering setting, traditional quantitative techniques are available for uncertainty quantification. Analysts can use these tools to measure uncertainty for single elements and to aggregate uncertainty across systems or systems of systems. Systems models are an important part of effective operation assessment at some levels, particularly for tactical assessment and terminal objectives. However, engineering models are inadequate to represent the full range of complex social and environmental aspects of operation assessment and program evaluation. Unfortunately, these difficult aspects are critical to successful operation assessment. While any model simplifies reality to enable timely decision making, the social-science models discussed here are better able to capture critical human aspects of warfighting. One trade-off for this increased capacity to represent complexity is that it is less straightforward to measure and communicate uncertainty for qualitative models.

The most effective operation assessments use both quantitative and qualitative approaches, also known as mixed methods. Next, we discuss approaches to addressing uncertainty for qualitative methods, quantitative methods, and mixed methods.

3.4.1 Qualitative Uncertainty

For causal strategic questions, there are three sources of structural uncertainty. First, *internal validity* addresses the degree to which the experiment demonstrates causation in addition to correlation (Cook et al., 2002). Second, *construct validity* addresses the degree to which the actual experiment reflects the intent of the strategic question (Cook et al., 2002). For instance, deterrence is a multi-faceted, complex

phenomenon that cannot be fully studied in a single assessment. While a strategic question might be interested in deterrence, an assessment study typically uses proxy measures to evaluate the level of deterrence. Third, *external validity* addresses the degree to which the results of an assessment of a specific instance generalize to other populations, operating environments, program implementations, and outcomes (Cook et al., 2002).

Uncertainty for these structural questions is typically both qualitative and subjective. However, it is a critical component of the overall uncertainty associated with the assessment. Assessors often address this type of qualitative uncertainty in the constraints, limitations, and assumptions of a study. Although this uncertainty is expressed qualitatively, it affects designs with either qualitative or quantitative data.

The most comprehensive way to communicate qualitative uncertainty is with a detailed narrative discussion addressing the complexity of the assessment environment and communicates the uncertainty in practical, decision-oriented terms. However, narrative reports are time consuming to read and may be at risk of being unused. One potential solution is to produce an executive summary highlighting the most important and unexpected aspects of the qualitative uncertainty. Another solution is to summarize the sources of uncertainty for each type of validity in a table and identify the most important contributors. The limitation of these summary methods is that they fail to convey important detail and may mask the true level of uncertainty. The assessment of qualitative uncertainty is necessarily subjective and is itself uncertain. For instance, the narrative description of the internal validity may describe the validity as high, when in fact there is an unexpected relationship that makes the validity low. Both the executive summary and the table summary entail an unavoidable loss of detail, but may be required to communicate results quickly. In the following subsections, we discuss internal validity, construct validity, and external validity in more

detail.

3.4.1.1 Internal Validity

Internal validity is the most basic form of qualitative uncertainty. To demonstrate full internal validity, assessors need to show the cause precedes the effect, there is an appropriate statistical relationship between the cause and effect, and no other factors explain the relationship between the cause and effect (Cook et al., 2002).

There are several sources of uncertainty that make it difficult to accurately evaluate and communicate the internal validity of an assessment. Even the relatively simple requirement of showing precedence can be difficult if only cross-sectional data are available (Cook et al., 2002). Cook et al. (2002) refer to the qualitative uncertainty associated with precedence as *ambiguous temporal precedence*. However, even in such cases, the assessor may be able to make pragmatic, documented assumptions to help establish plausible attribution in the absence of strong internal validity required for causation. One method to mitigate the limitations of cross-sectional data is to add a case study that illuminates the precedence relation between the suspected cause and effect (Cook et al., 2002). While this case study will necessarily be limited, it may provide more evidence to bolster the attribution claim.

The remaining sources of qualitative uncertainty for internal validity relate to the need to distinguish between the hypothesized cause and other causes. We address statistical uncertainty in Section 3.4.2.

Selection and *attrition* are qualitative sources of uncertainty related to the requirement that treatment and control groups are similar (Cook et al., 2002). The selection problem refers to the degree to which the treatment and control groups are different (Cook et al., 2002). In a descriptive or quasi-experimental design, assessors should assume that casual uncertainty due to selection exists (Cook et al., 2002). The

randomization used to create an experimental design addresses the selection problem probabilistically, but can still be inadequate with small sample size (Cook et al., 2002). Attrition is a subset of selection that deals with heterogeneous changes in the treatment and control groups over time (Cook et al., 2002). Over the course of an assessment, it is common for subjects to leave the original study groups. The change in group composition is also typically asymmetric between the control group and the treatment group, resulting in groups that are no longer similar (Cook et al., 2002). Pretesting in quasi-experimental designs helps to address both selection and attrition uncertainty (Cook et al., 2002). Reporting attrition rates can also provide an indication of the uncertainty associated with dissimilar experimental groups (Cook et al., 2002).

History, *maturation*, and *regression* relate to outside factors that also vary over time and which may be confused with the treatment effect (Cook et al., 2002). Historical uncertainty is the possibility that any other event that occurred between the intervention and data collection may have produced the observed outcome without the intervention occurring (Cook et al., 2002). This limitation is a fundamental problem for determining causation because there are an unlimited number of potential alternate explanations for any experiment. In practical application, assessors should attempt to insulate the test from outside influence as much as possible and explicitly address any remaining plausible outside factors that may be competing explanations. The major subjective distinction here is the difference between plausible outside explanations and unlikely explanations that are not worth considering. While assessors can address and reduce this type of uncertainty, they can never eliminate it completely.

Maturation refers to the problem of test subjects changing over time (Cook et al., 2002). Many stability operations take place over years, so the treatment and control

groups will change over time, possibly in different ways. This problem is addressed by selecting groups with similar circumstances, locations, and ages to minimize any difference in maturation (Cook et al., 2002). Regression also refers to changes in group characteristics over time, but is related to statistical fluctuation of the random noise in measurements (caused by both error and underlying factors) (Cook et al., 2002). In descriptive designs and quasi-experiments, subjects often enter the study with extreme measurements, either due to the intervention design or due to self-selection (Cook et al., 2002). Given the extreme measurement, following measurements are likely to be more typical, with or without the intervention (Cook et al., 2002). If the extreme selection is unavoidable, assessors can limit the negative effects by drawing both the treatment and control groups from the same extreme subjects, using multiple measures, or measuring at multiple points in time (Cook et al., 2002).

The assessment implementation can also affect the internal validity. *Testing* and *instrumentation* are ways the act of assessing can introduce uncertainty. Participation in the assessment, through surveys, interviews or data collection, may change the behavior of the subjects in a way that can be confused with the treatment (Cook et al., 2002). For example, if the design selects trainees as a control group for an existing training program, the additional data collection or observation may cause them to work harder and performer better than they would have performed without the observation. The instrumentation, or methods of observing and collecting data, may also change over time. Changes in instrumentation may cause changes in observation, even in the absence of the intervention (Cook et al., 2002). Data collectors should avoid changes in instrumentation but, if necessary, the collectors should use both the old and the new methods to allow for comparison (Cook et al., 2002).

All of the previous methods may occur independently or in an additive or interactive manner (Cook et al., 2002). Ideally, assessors can limit these sources of

uncertainty using randomization and appropriate sample sizes. In most practical cases, it will be necessary to investigate and respond to these issues using statistical techniques, design changes, and case studies.

In an applied military scenario, the instrumentation of military personnel and equipment during collective training events can increase training realism. For instance, instrumentation such as Multiple Integrated Laser Engagement System (MILES) simulates soldiers killed or wounded in action (N. Curtis, personal communication, July 10, 2018). However, uncertainty is introduced when the instrumentation loses power or malfunctions at some point during the training exercise (N. Curtis, personal communication, July 10, 2018). Unreliable instrumentation may also change soldiers' behavior and impact the manner in which they execute the mission. Observers often adjudicate as necessary and provide qualitative assessments of internal validity to augment the numeric data collection (N. Curtis, personal communication, July 10, 2018).

3.4.1.2 Construct Validity

Construct validity refers to the degree to which the actual experiment reflects the intended assessment question (Cook et al., 2002). Most causal assessment questions concern abstract concepts that are not directly measurable. If the measured aspects of the experiment do not adequately reflect the abstract concept, an experiment that has strong statistical and causal validity may still fail to provide strong evidence to answer an assessment question.

Consider the strategic question, "To what extent have group members been effectively removed by counter-network actions?" (Arnhart and King, 2018). Given a well-defined strategic question, the intent of the strategic question and the experimental design might diverge with respect to the populations, operating environments,

operational implementation, and outcomes (Cook et al., 2002).

In this case, the population under study is group members, but it is important that the decision-maker and assessment team are in agreement about the precise definition of group membership, beyond identifying the specific group. Does it include all self-proclaimed group members? Does it include group members that are geographically separated or inactive? Once the experiment has been conducted, this type of conceptual distinction can drive disagreement about the rigor or validity of the assessment results. The decision-maker and assessment team might agree to a restrictive definition of group members and only include some group members in the experiment. If auditing agencies interpret group members using a broader definition, their interpretation of the study results may differ. By making these types of abstract concepts explicit, assessors can reduce the potential for future confusion or disagreement.

Similar considerations apply to the other components of the assessment. For the operating environment, perhaps successful removal of group members in one geographic area increases recruitment in another geographic area. The operational implementation envisioned in the assessment and the actual implementation may also differ. The assessment may be framed with the assumption that the counter-network operations were conducted by well-trained operators while, in practice, the training levels of the operators may have been lower than expected. This risks attribution of negative results to counter-network operations rather than the level of training of the operators. Lastly, the conceptual outcomes are also vulnerable to differences between the concept and the concrete study. Effective removal may be approximated by lack of activity or contact with the group for a set period of time. However, if many of the discouraged group members rejoin the group after this period, the experiment would show an inaccurate improvement.

The task of making the concepts in the strategic question explicit is challenging. As in the case of ruling out alternative causal explanations, there are an unlimited number of interpretations of abstract concepts. In practice, the construct validity of an experiment will never be perfect, but the risk of divergence can be mitigated by addressing several common threats. We discuss the risks most relevant to military operations here, but Cook et al. (2002) provide a comprehensive list.

The primary threat to construct validity is an inadequate definition of the concepts under study, also called *inadequate explication of constructs* (Cook et al., 2002). Assessment teams should define concepts at the right level of specificity for the assessment and subsequent decisions. For instance, an assessment team might use the public perception of local governance as an indicator in an assessment of stability operations. Self-reported satisfaction and satisfaction elicited through other means may differ, so distinguishing between the two explicitly can reduce disagreement about results.

Assessors can also reduce uncertainty concerning the relationship between the experiment and the conceptual question by using multiple means of operational implementation and outcome measurement. This addresses the *mono-operation bias* and *monomethod bias* (Cook et al., 2002). If a high-level assessment question seeks to determine if key leader engagements are effective, an ideal assessment design would include several different types of key leader engagements. If the design is restricted to military-to-military engagements, the results may not reflect the effectiveness of all types of key leader engagements. Similarly, if an opinion-based outcome is measured only through in-person interviews, the concept measured may be publicly-acceptable opinion, while the addition of anonymous surveys might broaden the concept to include private opinions as well.

There are a broad class of threats to construct validity that result in an inability

to distinguish between the desired concept and the combination of the desired concept and the experiment. If operators, local populations, or even assessors are aware that an experiment or assessment is taking place, their behavior will reflect both the result of the operation and the awareness of the assessment (Cook et al., 2002).

Lastly, the risk of *treatment diffusion* refers to the possibility that the control group receives some or all of the effects from the treatment (Cook et al., 2002). A development project intended for a specific locality may directly or indirectly benefit nearby localities, so their status as a control group might be compromised. Assessors can mitigate the detrimental effects of treatment diffusion by measuring the degree to which each group is impacted by the treatment or operation, whether or not they are the intended target (Cook et al., 2002).

3.4.1.3 External Validity

External validity refers to the degree to which the experimental results generalize to other settings of interest (Cook et al., 2002). This generalization applies to each of the four components of the assessment: subject, operating environment, operational implementation, and outcome. While the generalization can include settings explored explicitly in the experiment, external validity is typically concerned with extrapolation outside the data collected in a single experiment (Cook et al., 2002).

If the assessment questions are narrow and relate only to the scenario under study, external validity may not be a concern. However, for ongoing operations, the data used for the assessment only cover a subset of all the cases of interest. In this scenario, a qualitative estimate of how well the results will transfer to other settings is important.

The major sources of uncertainty for external validity are the potential for unknown interactions between the observed causal relationship and each of the four

components of the assessment (Cook et al., 2002). Whether the subjects studied are people, military units, regions, or any other experimental group, the subjects not included in the experiment may react differently to the operation or program. Similar concerns hold for operating environment, operational implementation, and outcomes.

There is also the possibility of *context-dependent mediation*, in which the same causal relationship is found in different settings, but the exact process through which the effect is achieved differs (Cook et al., 2002). For example, security cooperation funding for a training program may cause increased effectiveness of partner nations in security efforts. However, it might be the case that one country used the funding to conduct more hours of training and another country used the funding to procure better training equipment. Although the cause and outcome are the same, the specific process differs. The best evidence for generalizability is to conduct a *meta-analysis* of multiple studies, either as a coordinated program or through historical review (Cook et al., 2002). When this type of evidence is unavailable, assessors can identify plausible interactions or barriers to generalization through professional experience or by comparing prior related work (Cook et al., 2002). By definition, such extrapolated external interactions will not have statistical evidence for or against the presence of interaction.

3.4.2 Quantitative Uncertainty

While qualitative uncertainty has a fundamental impact on the reliability and utility of an assessment, quantitative uncertainty is also critical to many decisions based on assessments. Despite the importance of both qualitative and quantitative uncertainty, qualitative uncertainty analysis is more common due to resource constraints, data limitations, and training practices. Furthermore, qualitative and quantitative certainty are not necessarily coupled: an assessment may have high statistical and

quantitative certainty, but low qualitative certainty.

Quantitative uncertainty includes both *aleatoric uncertainty*, caused by known variation, and *epistemic uncertainty*, caused by incomplete knowledge (Ferson et al., 2004). These categories are context dependent, but they can be useful for analyzing and communicating quantitative uncertainty. The decision making literature refers to these concepts as *risk* (Knight, 1921) when the probability distribution is known and *ambiguity* (Camerer and Weber, 1992) when the probability distribution is unknown. This distinction is also similar to the "known unknown" and "unknown unknown" popularized by former Secretary of Defense Donald Rumsfeld (Rumsfeld, 2002).

Methods for analyzing risk, or statistical variation, are mature and relatively well understood by decision-makers. Methods for analyzing ambiguity, however, are more complex to implement and less well-known. As is the case with quasi-experimentation in assessment design, there are approaches to partially address ambiguity when a full treatment is not feasible. Some modern data analysis techniques are powerful but lack associated methods for expressing quantitative uncertainty associated with the results. For instance, sentiment analysis packages typically do not include a standardized method of reporting the uncertainty associated with the calculated sentiment. It is important to address both the risk and the ambiguity associated with any assessment approach, using numerical and narrative approaches as necessary. The following subsections discuss statistical validity and uncertainty intervals in more detail.

3.4.2.1 Statistical Validity

In traditional statistical hypothesis testing, Type 1 error is the likelihood of a false positive, i.e., concluding that there is an effect when no effect exists. Type 2 error is the likelihood of a false negative, i.e., concluding that there is no effect when an effect exists. The complements of these errors are *confidence* and *power*, respectively.

A test with high confidence and high power has a low probability of reaching a false conclusion in either direction. Unfortunately, there is a trade-off between confidence, power, and experiment settings.

Low statistical power increases the uncertainty in quantitative results (Cook et al., 2002). The power of an assessment design depends on the confidence, the difference to detect, the noise in the measurements, and the sample size (Montgomery, 2017). Analysts typically determine the confidence level first, based on the expected costs of false positives. Confidence levels in practical designs may be lower than those commonly accepted in research, but if the confidence level is too low, the limited additional information the experiment provides may not justify the cost. The power of an assessment is sensitive to the difference the study is designed to detect. It is easier to test for large differences. Conversely, the more noise in the measurement, the lower the power of the design. Analysts typically interview stakeholders to determine a meaningful detection difference. Is a one-percent decrease in violence operationally meaningful? Is a ten-percent reduction meaningful? Measurement noise, on the other hand, is intrinsic to the method of measurement. Lastly, analysts choose sample size to balance power and cost. The more samples collected, the higher the power of the experiment. There are many other ways to increase the power of an experimental design (e.g., Cook et al., 2002).

Restriction of range can also reduce the power of a statistical test (Cook et al., 2002). To increase power, analysts should design tests to include the conditions which they expect to produce the largest difference in effect between groups. This design typically involves selecting conditions that are near the extremes of the possible values of a variable under study, while staying within normal operating limits. For the same reasons, designs should use continuous scales whenever possible for quantitative variables. For example, the raw miss distance of a weapon impact is preferable to

categorizing impacts as hit or miss.

Violated assumptions of the test statistics also increase uncertainty about the results of a statistical test, but in a less direct way (Cook et al., 2002). However, analysts can usually evaluate statistical assumptions with both subjective expertise and formal tests of assumptions (e.g., normality tests, correlation). Some tests are robust to violations of assumptions while others are not.

The *error rate problem* exists when multiple statistical tests are conducted. The probability of error in a single test is lower than the combined probability of error in a group of tests because the probability of at least one error compounds. For formal hypothesis tests, the Bonferroni correction provides a conservative method to account for multiple tests (Cook et al., 2002). *Fishing* is another version of the error rate problem that exposes the analysis to the risk of spurious statistical relationships if enough relationships are tested (Cook et al., 2002).

Undesirable variability in the operational implementation, the operating environment, or the subjects can introduce quantitative noise that obscures the effect of interest (Cook et al., 2002). Differences in operational implementation for different data points may or may not be desirable. In some cases, the design may tailor the implementation to different experimental subjects or operating environments (Cook et al., 2002). However, if the quality of the implementation varies undesirably, it reduces the power of the test (Cook et al., 2002).

Assessments can never control the operating environment to the same degree that academic or field studies control their environments. To address this limitation, assessors can use quantitative measures and qualitative descriptions of the actual operational environment during implementation to adjust the analysis to account for changes (Cook et al., 2002). Variability in subjects increases noise, but it can also provide more evidence for generalization (Cook et al., 2002). If the noise associated

with variability in subjects is too high, one solution is to *block* the design against the different types of subjects (Montgomery, 2017). This technique measures the variability in subjects separately from the random variations due to noise.

An alternative to presenting a formal hypothesis test is to report effect size, confidence intervals, confidence, and power (Cook et al., 2002). This avoids the artificial distinction between "statistically significant" and "statistically insignificant," while also ensuring a pragmatic treatment of effect size and power (Cook et al., 2002). In some cases, a result may be statistically significant but practically insignificant. In other words, the analysis might show with high confidence a very small difference between two groups that has no impact on operations. Conversely, statistically weak evidence of a potentially large difference in groups might justify a follow-on study.

3.4.2.2 Uncertainty Intervals

Whenever possible, assessments should avoid point estimates because they imply an implausible level of precision. Interval estimates help to represent the quantitative uncertainty in the results more accurately. However, interval estimates are vulnerable to the same type of criticism as point estimates: the exact bounds of the interval are not known precisely and are themselves subject to second-order uncertainty. But the second-order uncertainty is also subject to third-order uncertainty and so on. In principle, we can model arbitrarily higher orders of uncertainty, but the additional modeling complexity quickly becomes intractable. Uncertainty models higher than second-order models are rare.

In time-critical scenarios, a point estimate may be the best option, preferably with a qualitative explanation of the remaining uncertainty. At the other extreme, second-order uncertainty, represented by a set of probability intervals, is appropriate when there is sufficient time for a detailed analysis and follow-on decisions rely critically on

a high-fidelity model of the ambiguity. In most operation assessment scenarios, however, a first-order uncertainty model using an interval estimate provides a reasonable trade-off between fidelity, speed, and complexity. Again, assessors can mitigate the remaining uncertainty by addressing it qualitatively in narrative form.

Within the class of interval uncertainty representations, there are several different options. The most familiar option is the *confidence interval*. In practical terms, confidence intervals define an interval that usually contains a parameter of interest. Strictly speaking, from a frequentist point of view, a $c\%$ confidence interval is defined so that if confidence intervals were calculated from an arbitrarily large number of independent samples, the proportion that include the true mean approaches $c\%$ (Vardeman, 1992). The formal distinction has little bearing on practice. Furthermore, Bayesians take the explicit view of subjective probability, using *credible intervals* for the same intent but with the strict interpretation of the interval having a $c\%$ subjective probability of containing the parameter. We can also define confidence intervals for other parameters that might be of interest (e.g., median, variance).

Prediction intervals are defined similarly, but instead of creating an interval around a parameter, the interval is defined around the next predicted value (Vardeman, 1992). This type of interval is appropriate when the follow-on decision depends on the predicted value of a single outcome, rather than the mean value. The prediction interval will be larger than the confidence interval. Furthermore, as the sample size increases, the size of the confidence interval decreases toward zero, but the size of the prediction interval approaches the non-zero, true variability of the process (Vardeman, 1992). In other words, the prediction interval includes the error due to sampling and the inherent variability of the process that generates the random variables. Prediction intervals can be interpreted from either a frequentist or Bayesian perspective, and they can also be calculated with parametric or non-parametric methods, which make

fewer or no distributional assumptions.

Tolerance intervals are somewhat more complex, but are often useful to decision making. A tolerance interval contains at least p proportion of the population with at least $c\%$ confidence (Hahn and Meeker, 2011). For example, the 95% tolerance interval for 99% of the population gives the interval that contains at least 99% of the population with 95% confidence. This type of interval provides statistical guarantees for the value of groups of observations. In many cases, the tolerance interval is more relevant to decision making than the relatively narrowly focused confidence interval. As with the other intervals, there are both frequentist and Bayesian interpretations and parametric and non-parametric methods.

Statistical intervals provide the best representation of interval uncertainty but, if necessary, descriptive intervals can also give an indication of uncertainty. These descriptive intervals are defined by percentiles, for example the 2.5% percentile and the 97.5% percentile. Quartiles, used in box plots, and the minimum and maximum values are special cases of sample percentiles. Although the descriptive interval is a simple and fast way to give an indication of variability, it underrepresents the uncertainty due to sampling error (Hahn and Meeker, 2011). The communication of assessments with these intervals should include an appropriate qualitative explanation of their limitations. For further details, a short introduction to statistical intervals is available in the work by Vardeman (1992) and a full treatment is available in the work by Hahn and Meeker (2011).

3.4.3 Mixed Methods

As in the field of operation assessment (Mushen and Schroden, 2014), the broader evaluation literature discusses the relative merits of qualitative and quantitative methods (Fitzpatrick et al., 2011). However, the modern consensus in program evaluation

is that a mix of both quantitative and qualitative methods tailored to the evaluation questions is the most effective approach (Fitzpatrick et al., 2011; Wholey et al., 2010). Often, the mixed-method approach includes several different qualitative and quantitative methods for a single evaluation. Mushen and Schroden (2014) arrive at a similar conclusion for operation assessment, although with more skepticism for quantitative methods. Military assessments are conducted at a variety of levels for different purposes. Accordingly, the appropriate mix of quantitative and qualitative methods varies with the purpose.

Mixed methods support five main purposes: triangulation, complementarity, development, initiation, and value diversity (Mathison, 2005). *Triangulation* uses several different proxy measures for a single abstract concept, with the intent of broader coverage of the concept (Fitzpatrick et al., 2011). Ideally, the evidence from the set of proxy measures converges in a coherent way, providing stronger evidence for answering the assessment question. *Complementarity* has a similar motivation but seeks out different results to highlight the complexity of the concept (Fitzpatrick et al., 2011).

The *development* and *initiation* purposes are ways of employing additional methods to improve follow-on assessments. The development approach uses additional testing (e.g., case studies, cross-sectional data) to refine the measures (Fitzpatrick et al., 2011). Initiation uses additional methods to find unexpected results that may require follow-on assessments to move in an entirely new direction (Fitzpatrick et al., 2011).

Mixed methods also address *value diversity* across stakeholders in the assessment (Fitzpatrick et al., 2011). This solution is a useful way to address the different needs of line of effort (LOE) owners, senior-decision makers, and oversight organizations. The best method to produce data useful for LOE owners will not be the same as the best method to satisfy summative accountability questions. Instead of choosing a single

compromised method, the mixed method approach suggests choosing a combination of methods that satisfy all stakeholders. The additional methods will likely enhance and complement each other, even for the stakeholders who are not the primary consumers for a given method. However, cost restrictions will likely limit the extent to which the set of methods can be tailored to every stakeholder.

Given the mix of quantitative and qualitative methods, assessors need to produce a report that synthesizes the qualitative and quantitative uncertainty from the analysis of structured and unstructured data. Generally speaking, assessments can address the qualitative and quantitative uncertainty separately, but it is important to avoid false certainty in the quantitative uncertainty reporting if there is high qualitative uncertainty in the supporting model. For instance, a confidence interval might be tight, but if the internal validity is weak, the assessment should include a clear caveat associated with the precise numeric result.

3.5 Conclusion

In this chapter, we highlighted the significant overlap between the practice of military assessment and program evaluation. The framework for identifying, analyzing, and reporting uncertainty in an evaluation context has value for the military assessment context. Evaluation theory provides a broad range of assessment paradigms for different contexts, which can be tailored to a particular staff structure or organizational level. Given an assessment paradigm, evaluation and monitoring also provide different levels of rigor and cost for the assessment design and data collection. Assessors can use this set of options to determine if a descriptive, quasi-experimental, or experimental approach is most appropriate. Then, the assessment report should include appropriate analysis of the qualitative and quantitative uncertainty for the selected design. A mixed-method approach that incorporates both quantitative and

qualitative methods produces the highest-quality assessment.

The assessment designs and uncertainty framework discussed here serve two primary purposes. First, they can help assessors improve assessment conduct. The complex implementation environment will always require modification to the basic forms of evaluation presented here, but they can serve as an initial assessment design. The second purpose for using program evaluation and performance monitoring as a foundation for assessment is that assessors can use them to rigorously justify methodology choices to oversight organizations.

As the military assessment community continues to professionalize and work toward improved assessment, there is opportunity for further research at the intersection of military assessment and program evaluation. Empirical studies and meta-analysis of previous assessments could provide concrete evidence that explores the frequency of different types of strategic questions under consideration, what types of assessment designs were used to answer those strategic questions, and how successfully those designs answered the strategic questions. Military assessors conduct evaluation under unique conditions, including a dynamic environment, adversarial setting, and shifting political objectives. More research on best practices for assessment design and uncertainty modeling under these conditions could also help to grow the set of tools available to practitioners.

IV. Robust Queue Inference Under Uncertainty

4.1 Introduction

Although queues have been well studied, much of the prior work has focused on the performance of a queueing system given its internal structure, arrival rate, and service rate (Asanjarani et al., 2017). In this analysis, we consider the opposite perspective: given a queue's input and output, we infer conclusions about its internal structure.

4.1.1 Problem Description

In particular, this research is related to adversarial applications in both military and commercial settings. One of the primary research areas with potential applications for unobservable queue inference is cybersecurity, including security games for modeling cyber attack and defense. In this application area, attackers use malicious software to degrade or infiltrate a defender's computer network. Attackers' actions are commonly modeled using attack trees, which often assume attacker knowledge of defender nodes (Mauw and Oostdijk, 2005). For instance, in standard game theoretic cybersecurity formulations, the attacker requires full or partial knowledge of the number of nodes in the defender's system to develop their optimal policy. Kiennert et al. (2018) provide examples of this modeling approach in cybersecurity resource allocation games (e.g., Otrok et al. (2008) and Chen and Leneutre (2009)) and in countermeasure optimization games (e.g., Lye and Wing (2005) and Shen et al. (2007)). To meet the assumptions for these game theoretic models and other attack tree models, the attacker must know the number of nodes in the defender's system. One way for the attacker to collect this information is to use surveillance of traffic outside the closed system to infer the hidden number of servers. Knowledge of the number of nodes in a closed system allows the attacker to develop more accurate and

effective strategies. Accordingly, it is valuable for the defender to know how much information the attacker can derive about the number of defended nodes using only unsecured traffic.

The problem of unobservable queue inference is closely related to terror plot identification and interdiction. Terror queues, introduced by Kaplan (2010) and extended by Seidl et al. (2016), model terrorist plot development and interdiction as a queue, where terror plots are customers and intelligence agents are servers. From an adversarial perspective, it is valuable for the terrorist organization to infer the number of intelligence agents available without access to internal information of the security organization. This problem is also relevant to generalized border security and adversarial interdiction problems that maintain a similar structure but are outside the specific application of terrorism.

This work also has applications in the broader commercial setting. Park et al. (2011) identify several potential application areas for unobservable queue inference, including internet traffic analysis, estimating a competitor's production capacity and queueing delays, and working with internal but incomplete historical data. Similarly, Guo et al. (2016) present an application of service time characterization under incomplete data in the setting of hospital emergency departments. In all of these application areas, real arrival and departure time data may include noise or may be limited to small samples, but current estimation methods are designed for exact data with large samples (Park et al., 2011).

In this context, we consider a multi-server, single queue with independent arrivals and no restriction on arrival distribution or service distribution. Using standard notation, this is a GI/G/c queue (Gross et al., 2011). The GI/G/c queue also implies unlimited queue length and an infinite pool of customers. We assume that we cannot observe any parameters internal to the queue, including queue length, service times,

number of servers, and service rates. However, we assume that we can observe arrival times, departure times, and customer identity so we can match the arrival and departure times of a given customer. We also consider a relaxation of perfect observation of these data sources. The goal is to determine all unknown parameters of the queue. Once we have determined the number of servers, we can determine all other quantities of interest (e.g., wait in queue, number in queue, service length, number in service) and estimate the corresponding parameters (Park et al., 2011). This analysis focuses on estimating the number of servers in a GI/G/c queue from arrival order, departure order, and customer identity.

4.1.2 Related Work

Queue inference was originally used to refer to estimation of queue parameters, such as queue length and wait in queue, from transaction data (Larson, 1990; Bertsimas and Servi, 1992; Jones and Larson, 1995). Here, we use it in a more general sense to refer to queue parameter estimation when the service is unobservable. Our work is most closely related to Park et al. (2011) who proposed the variance minimization method to address this problem. We take their work to be the standard approach and use it as a baseline to compare performance.

The developments presented in this chapter are tangentially related to the work of Dong and Whitt (2015), which involves analyzing birth-death processes, but their focus is on the limiting distribution of the number of customers in the system for arbitrary birth-death processes. Our work is also related to the research by Frey and Kaplan (2010), who consider a similar problem using only periodic departure data but make additional distributional assumptions by assuming Poisson arrivals. Their results are also subject to the limitations described by Jones (2012).

Unobservable and partially unobservable queues are also considered in the litera-

ture dealing with economic models of balking and reneging (Edelson and Hilderbrand, 1975; Jones, 1999; Burnetas and Economou, 2007; Pazgal and Radas, 2008). This area of research is conducted from the customers' perspective with the goal of making decisions about joining different types of queues, rather than estimating the number of servers. This remains an active area of research (Yu et al., 2016; Guha et al., 2016), but to our knowledge Park et al. (2011) published the benchmark in server estimation for GI/G/c queues with unobservable service.

We present the methodology for our new approach and the existing approach in Section 4.2. We describe the experimental design in Section 4.3 and discuss results for small samples and noisy observations in Section 4.4. We extend the primary results to address the non-preemptive last-come, first-served (LCFS) case in Section 4.5. Finally, we conclude with a summary and recommendations for future work in Section 4.6. Supporting theory is presented in Appendix A Section 1.1 and Appendix A Section 1.2 and details for accessing code and data are provided in Appendix A Section 1.3.

4.2 Methodology

This section presents an order-based method for estimating the number of servers in an unobservable queue. The stable order of a first-come, first-served (FCFS) queue prior to service presents an opportunity for an efficient server estimation method. We compare this new approach to the existing variance minimization approach. We also show that a modified extension of the order-based method to LCFS queues has similar convergence performance to previous techniques, but may display greater error prior to convergence in some cases. To implement the analysis methods, we use the Julia computer programming language (Version 0.6) (Bezanson et al., 2017) with the SimJulia, DataFrames, Distributions, StatsBase and BenchmarkTools packages

(Lauwens, 2018; Arslan et al., 2018; JuliaStats, 2018a,b; Revels, 2018). First, we build a standard GI/G/c queue simulation, then we apply each algorithm to the output of the simulation to compare performance, using JMP for statistical analysis (SAS Institute, 2016).

4.2.1 Variance Minimization Method

The variance minimization method is an existing approach proposed by Park et al. (2011). This method can be used for both FCFS and LCFS queues with minor variations. We focus primarily on FCFS queues, but a discussion of the LCFS case is included in Section 4.5. Park et al. (2011) prove that the variance of estimated service times achieves a minimum at the actual number of servers. Using this theorem, they take the solution of the minimization problem in Eq. (20) as the estimate for the number of servers in the queue. Table 4 defines the parameters and variables used in Eq. (20) (Park et al., 2011). Note that we replace $\hat{c}$ with $\hat{c}_n$ to emphasize that the estimator depends on the number of observations.

$$\underset{\hat{c}_n}{\text{minimize}} \quad \sum_{i=1}^{n} \left(\hat{S}_i - \bar{\hat{S}}\right)^2 \tag{20a}$$

$$\text{subject to} \quad \hat{B}_i = A_i, \qquad i = 1, \ldots, \hat{c}_n, \tag{20b}$$

$$\hat{B}_i = \max\left(A_i, D_{(i-\hat{c}_n, i-1)}\right), \quad i = \hat{c}_n + 1, \ldots, n, \tag{20c}$$

$$\hat{S}_i = D_i - \hat{B}_i, \qquad i = 1, \ldots, n, \tag{20d}$$

$$\hat{S}_i > 0, \qquad i = 1, \ldots, n, \tag{20e}$$

$$1 \leq \hat{c}_n \leq n, \tag{20f}$$

$$\hat{c}_n \in \mathbb{Z}. \tag{20g}$$

The objective function (20a) represents the sum of squared error of the estimated service times and the estimated mean service time. The decision variable is the

Table 4. Parameter and Variable definitions

Variable	Definition
A_i	Arrival time of the ith arrival
X_i	Interarrival time following the ith arrival
S_i	Service time of the ith arrival
D_i	Departure time of the ith arrival
$D_{(k,m)}$	kth order statistic among the first m departure times
c	Number of servers
n	Total number of customers
$\tilde{c}_j$	Running estimate for the number of servers after jth departure
$\hat{c}_j$	Cumulative estimate for the number of servers after jth departure

estimated number of servers. Constraint (20b) sets the estimated begin-service time to the actual arrival time, for the first $\hat{c}_n$ customers. For the subsequent customers, constraint (20c) sets the ith estimated begin-service time to either the ith arrival time (corresponding to a customer arriving to an empty queue) or to the departure time of the $(i - \hat{c}_n)$th departure (corresponding to a customer waiting in queue then taking the first available server). Constraint (20d) sets the estimated service time to the difference between the true departure time and the estimated begin-service time. Constraints (20e), (20f), and (20g) ensure estimated service times are positive and that the estimated number of servers is feasible.

We observe that this optimization problem is equivalent to minimizing the sample variance of estimated service times. However, to accurately use the theoretical variance minimization results, we need sufficient data for the sample variance to converge to the population variance. In the general service case with unknown service distributions, this requirement is non-trivial. Park et al. (2011) show good results for samples of $n = 1000$ with exact arrival and departure times. In this application area, we are interested in good performance for small, noisy samples in addition to the long-run performance.

4.2.2 Order-based Method

The approach presented here is a function of arrival and departure order rather than arrival and departure times. Due to these reduced information requirements, we refer to it as the order-based method. This approach is motivated by the straightforward observation that the number of servers can be estimated directly by considering the number of customers that are in service. For an FCFS queue, given the ith customer to arrive has departed the system, all $i-1$ customers that arrived before that customer must have already entered service. Of those $i-1$ customers, any customers that have not yet departed must still be in service. The number of customers in service can be used to identify a lower bound on the number of servers. Although the order-based approach is straightforward, the theoretical and experimental results in the following sections suggest that this method has the best published performance for unobservable queue inference with small, noisy samples.

The methodology for this approach is formalized in Algorithm 1, which uses set operations to parallel the logic of the algorithm development discussed in the previous paragraph. Algorithm 2 provides a fast implementation that yields the same results as Algorithm 1, with variable and function definitions for both algorithms shown in Table 5. The main change in Algorithm 2 is that, instead of using a mathematical set difference operator, which is computationally expensive, we manually track the estimated number of servers occupied after each departure with incremental updates. Note that, depending on the arrival index of a departing customer, it is possible to update the estimated number of servers by more than one, as shown in the conditional if-statement of Algorithm 2. In these algorithms and the remainder of the chapter, we use $X[i]$ to refer to the ith component of the vector X. For instance, *DepartOrder[3]* is the third component of the *DepartOrder* vector.

The set-based server estimation algorithm has worst-case complexity of $O(cn^2)$

Table 5. Variable and function definitions for Algorithm 1 and Algorithm 2

Name	Type	Definition
DepartOrder	Vector of integers	Arrival indices of first n departed customers, sorted by departure order
n	Integer	Number of departed customers
MaxCustomer	Integer	Largest arrival index of first i departed customers
NumServers	Integer	Estimated number of servers
CustIndexAll	Set of integers	Set of arrival indices of customers known to have entered service given the first i departed customers
CustIndexDeparted	Set of integers	Set of arrival indices of first i departed customers
CustIndexInService	Set of integers	Set of arrival indices of customers known to be in service given the first i departed customers
NumServersOccupied	Integer	Estimated number of servers occupied
MaxCustomerNext	Integer	Largest arrival index of first i departed customers and the previous *MaxCustomer* value
Max	Function	Maximum of two integers
MakeSet	Function	Convert vector of integers to set of integers
SetDifference	Function	Return the elements of the first set that are not in the second set
Length	Function	Return the number of elements of a vector

where c is the number of servers and n is the number of observations (departed customers). The complexity is calculated from n loops of a set-difference operation between sets of size n and cn in the worst case, which gives an inner-loop complexity of $O(n+cn)$. The order-based algorithm, shown in Algorithm 2, gives the same results with $O(n)$ complexity. Algorithm 2 is used for all data collection and throughout this chapter as the order-based algorithm. The variance method uses a mixed integer non-linear programming formulation, but can be solved directly with a brute force search over a small search space (Park et al., 2011). Micro-benchmarks on a 2.4 GHz Intel Core i5 CPU with 8 GB RAM show a substantial relative reduction in computational time for the order-based method. For a 7 server queue with 1,000 customers, the variance method has a median computation time of 225 ms with a minimum of 218 ms and a maximum of 231 ms over 23 samples. For the same queue, the order-based method has a median computation time of 0.462 ms with a minimum of 0.444 ms and a maximum of 6.101 ms over 8,709 samples. Note that the number of samples is tuned to the computation time, resulting in a large sample size for the order-based benchmark. Although the order-based method has a large relative reduction, both methods are fast enough for the projected application areas. The source code, testing, benchmarks, and data are available online, with details in Appendix A Section 1.3.

The numerical example in Table 6 illustrates the algorithm for a small sample. The running estimate corresponds to the estimate calculated in the inner loop of the algorithm while the cumulative estimate corresponds to the final estimate, for a given sample size. Jobs are listed in departure order since the estimates are updated upon each departure event. In the notation that follows, $\tilde{c}_m$ is used for the running estimate produced after the mth departure, and $\hat{c}_m = \max_{k\in\{1,\ldots,m\}} \tilde{c}_k$ is used for the mth cumulative estimate, which is equivalent to the estimate produced by the order-based method.

Algorithm 1 Set-based Server Estimation

```
1: procedure SetEstimator(DepartOrder)
2:     n ← number of departed customers
3:     MaxCustomer ← 0
4:     NumServers ← 0
5:     for i = 1 : n do
6:         MaxCustomer ← Max(DepartOrder[i], MaxCustomer)
7:         CustIndexAll ← MakeSet(1 : MaxCust)
8:         CustIndexDeparted ← MakeSet(DepartOrder[1:i])
9:         CustIndexInService ← SetDifference(CustIndexAll, CustIndexDeparted)
10:        NumServers ← Max(Length(CustIndexInService) + 1, NumServers)
11:    end for
12: return NumServers
13: end procedure
```

Algorithm 2 Order-based Server Estimation

```
procedure OrderEstimator(DepartOrder)
    n ← number of departed customers
    MaxCustomer ← 0
    NumServers ← 0
    NumServersOccupied ← 0
    for i = 1 : n do
        MaxCustomerNext ← Max(DepartOrder[i], MaxCustomer)
        if DepartOrder[i] < MaxCustomer then
            NumServersOccupied ← NumServersOccupied − 1
        else
            NumServersOccupied ← NumServersOccupied + MaxCustomerNext −
MaxCustomer − 1
        end if
        MaxCustomer ← MaxCustomerNext
        NumServers ← Max(NumServers, NumServersOccupied + 1)
    end for
return NumServers
end procedure
```

Table 6. Order-based algorithm example

Customer ID	Arrival Index	Departure Index	Running Estimate	Cumulative Estimate
79305	1	1	1	1
89698	3	2	2	2
48692	5	3	3	3
87269	4	4	2	3
21094	2	5	1	3
26318	6	6	1	3

The order-based method is subject to the mild restrictions listed in Table 7, where traffic intensity is denoted by ρ. These restrictions are developed from the discussions by Kiefer and Wolfowitz (1955), Whitt (1972), and Kennedy (1972). Assumption (iii) is met for any pair of independent distributions with support such that $\max(\text{support}(X_i)) > \min(\text{support}(S_i))$, where X_i is the interarrival distribution and S_i is the service distribution. Note that Assumptions (iii)-(vi) are met if interarrival and service distributions are independent and continuous with support over $[0, \infty)$. Appendix A Section 1.2 provides more detail on the motivation for each assumption. The variance method implicitly assumes the first two criteria, but does not explicitly require the remaining criteria. In particular, the variance method is appropriate for deterministic service distributions, as discussed further in Section 4.4.

Table 7. Assumptions

Index	Assumption
(i)	Customers are uniquely identifiable
(ii)	$\rho < 1$
(iii)	$P(X_i > S_i) > 0$
(iv)	Support of interarrival distribution includes 0
(v)	Interarrival and service distributions are absolutely continuous
(vi)	Interarrival and service distributions are independent

Given the assumptions in Table 7, we prove three FCFS results.

Proposition 2. *Given an FCFS GI/G/c queue that meets Assumptions (i)-(vi),*

there is a positive probability that the order-based estimation algorithm produces the correct estimate immediately following the cth departure.

Proposition 3. *Given an FCFS $GI/G/c$ queue that meets Assumptions (i)-(vi), the order-based estimation algorithm produces an estimate that is a lower bound on the true number of servers.*

Theorem 4. *For an FCFS $GI/G/c$ queue that meets Assumptions (i)-(vi), the order-based server estimation algorithm produces estimates $\hat{c}_n$ such that* $\lim_{n\to\infty} P(|\hat{c}_n - c| > \epsilon) = 0 \quad \forall\, \epsilon > 0$.

The proofs for these results are presented in Appendix A Section 1.1. Proposition 3 shows that the order-based approach provides a lower bound on the true number of servers, regardless of sample size. This guarantee may be useful to decision makers in practical application. Proposition 2 and Theorem 4 show that the order-based method produces an estimate that converges in probability to the true number of servers. These results provide a theoretical justification for using the order-based method for arrival and service distributions that are not considered in the experimental results.

4.3 Experimental Design

Given the theoretical results in Section 4.2 and Appendix A Section 1.1, we now consider a simulation that explores the performance of the order-based method across several factors of interest. We first consider the standard scenario in which arrival and departure times are observable without error and characterize the performance of the order-based approach relative to the baseline of the variance approach. Then, we relax the precise observation assumption by introducing error into the arrival and departure times to evaluate the robustness of the order-based approach. The LCFS setting is considered separately in Section 4.5.

We are interested in two responses: average estimation error and the number of observations required to approximately converge. The ideal estimator converges quickly to the true value and has low error before it converges. In both cases, we introduce approximations to address computational limitations and to explore a larger design space. First, we define average estimation error as $\frac{1}{k}\sum_{i=1}^{k} \mid c - \hat{c}_i \mid$. In particular, we consider a 50 observation window, then take the average of the absolute difference between the true number of servers and the estimated number of servers. The estimation error from a particular number of observations on a particular experimental run is relatively noisy. By taking the average estimation error over a window, the response has less noise at a cost of lower fidelity. This trade-off reduces the size of the experiment. The numerical setting of 50 observations for the window size was selected after pilot runs of the simulation, which were used to balance coarseness and computation time.

The second response is the number of observations needed to approximately converge. Given the finite nature of any simulation, we can only report approximate convergence. We define approximate convergence as zero estimation error for 200 consecutive observations and calculate it by sampling every 20 observations, with a maximum of 1000 total observations. Again, this definition balances coarseness against experiment size. The particular numerical settings for this sampling rate were set after initial pilot runs.

We consider six factors in the design space: estimation method, number of observations, number of servers, traffic intensity, arrival distribution, and service distribution. These factors are summarized in Table 8. The primary factor of interest is the estimation method. The number of observations refers to the number of customers that have been processed through the queue and departed. The ideal estimator would perform well even with a low number of customers observed. Note that this factor

is ignored for the second response, number of observations to converge. That is, we calculate the error at the set number of observations, then let the simulation run until approximate convergence or the maximum observation limit is reached. The observation and server factors are modeled as continuous factors although they are technically discrete.

The remaining factors are included to characterize the performance of the estimators over different environmental conditions. The ideal estimator would be robust to changes in all of these factors. The traffic intensity is set to maintain comparability between the different settings for the GI/G/c queues. Although the theoretical work only requires traffic intensity in $(0, 1)$, we set the levels to represent queues that are not over-served. The service distributions were chosen to explore the breadth of possibilities in a GI/G/c queue. The exponential distribution is a classic service distribution for queueing models and serves as a baseline. The log-normal distribution has higher variance than the exponential distribution and is also explored by Park et al. (2011). We also test two distributions with finite support. The uniform distribution has a substantially different shape than the other three distributions while the beta distribution has a similar shape to the exponential distribution but has finite support. This mix of characteristics allows us to explore the performance of the estimator under a variety of conditions.

Initial test runs for the simulation indicated that the arrival distribution did not have a large effect on performance. To reduce the size of the experiment, the arrival distribution was limited to two levels: the baseline exponential case and the least similar case, the uniform distribution. The observations, servers, and traffic intensity were set to three levels because initial test runs indicated non-linearity.

We developed a full factorial experimental design with both estimation methods applied to each run, which results in 216 runs per replication. We conducted four total

replications to achieve the desired precision for comparison. We set the confidence level to 0.95 resulting in a power above 0.90 for all three-way interactions and above 0.97 for all lower-order terms.

Table 8. Factors

Name	Type	Levels
Estimation Method	Categorical	Order-based, Variance
Observations	Numeric	40, 200, 400
Servers	Numeric	2, 9, 15
Traffic Intensity	Numeric	0.70, 0.90, 0.99
Arrival Distribution	Categorical	Exponential, Uniform
Service Distribution	Categorical	Exponential, Log-Normal, Uniform, Beta

4.4 Results

4.4.1 Estimation Error

The estimation error data collected in the experiment exhibit a disproportionate number of zero values, as shown in Figure 10. The two-server runs and high observation runs often had no estimation error. To address this non-normality in the data, we fit a zero inflated Poisson regression model (Lambert, 1992). Although the estimation error is not strictly count data, as it is averaged over a window of observations, the departure is not severe (Gourieroux et al., 1984). We explored a two-stage model and a neural network as confirmatory models and found similar results.

Overall, applying a t-test, we find a statistically significant difference in the average estimation error with a p-value less than 0.001. The mean estimation error for the variance method is 0.81 servers, while the mean estimation error for the order-based method is 0.27 servers, a 66% reduction in error. However, the regression model provides evidence for multiple higher order interactions that are statistically significant, which indicates that the relative performance of the order-based method is not constant as other factors change. In some cases, the reduction in error is three

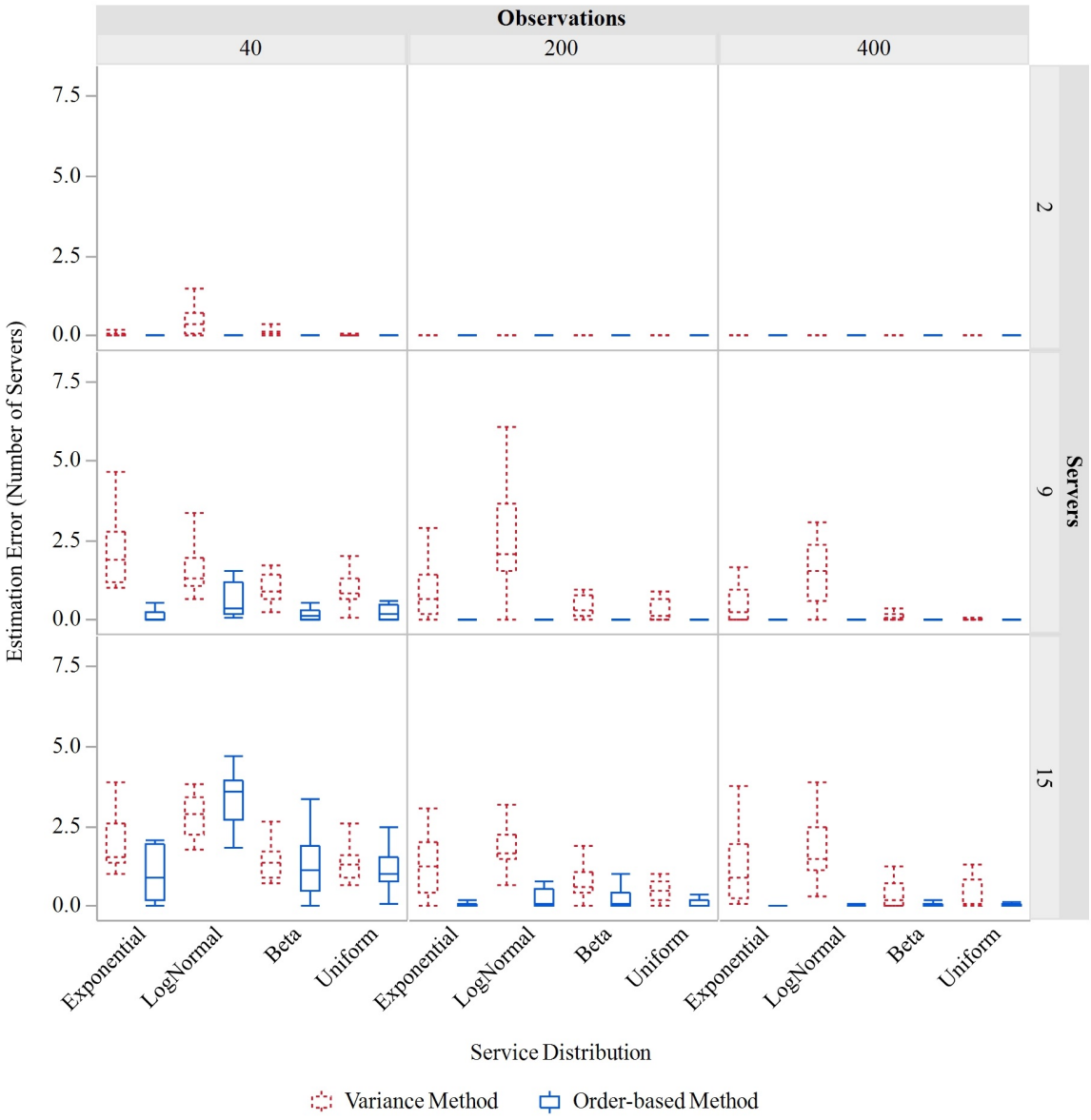

Figure 10. Average estimation error

servers or higher. All model terms with p-value less than 0.0001 are shown in Table 9. The generalized R-squared is 0.63 on a validation set. Since this is a computer simulation, all sources of variance are controlled. These results suggest that much of the variance in estimation error is due to the stochastic nature of the interarrival and service times.

Table 9. Significant model terms

Term	p-value	Coefficient
Estimation Method[Variance Method]	< 0.0001	3.19
Observations	< 0.0001	-0.01
Servers	< 0.0001	0.55
Traffic Intensity	< 0.0001	-6.34
Estimation Method*Observations	< 0.0001	0.01
Estimation Method*Servers	< 0.0001	-0.29
Estimation Method*Traffic Intensity	< 0.0001	3.37
Observations*Traffic Intensity	< 0.0001	-0.01
Servers*Servers	< 0.0001	-0.03

On average, the regression model estimates that the order-based method produces a server estimate with error that is 3.19 servers lower, with a 95% confidence interval of (2.26, 4.11). However, due to the higher-order interactions, this difference is not constant across factor settings. In many cases the order-based method performs much better than the variance method. Under nominal conditions of exponential arrivals and traffic intensity of 0.9, Table 10 shows the model predictions for each combination of factors, where the asterisk indicates a non-statistically significant difference using individual confidence intervals. There is a statistically significant improvement for all but three factor setting combinations. In two of those cases, there is a non-statistically significant degradation in performance. While this degraded performance is not statistically significant, the performance of the order-based method and variance method is not clearly different for the combination of low observations and high servers in the absence of measurement error. However, when there is measurement error, the order-based method shows a clear improvement even in the low observation, high

server cases, as shown in Figure 13.

Table 10. Model predictions

Factor			Order-based Method		Variance Method		Result
Obs.	Serv.	Service	Mean	95% Conf. Int.	Mean	95% Conf. Int.	% Impr.
40	2	Expo	0.001	(0.000, 0.003)	0.122	(0.069, 0.214)	99
40	2	LogNorm	0.008	(0.003, 0.019)	0.363	(0.185, 0.712)	98
40	2	Beta	0.001	(0.000, 0.002)	0.041	(0.021, 0.078)	98
40	2	Unif	0.001	(0.000, 0.003)	0.036	(0.018, 0.071)	97
40	9	Expo	0.094	(0.050, 0.177)	1.713	(1.297, 2.263)	95
40	9	LogNorm	0.601	(0.411, 0.880)	3.490	(2.504, 4.865)	83
40	9	Beta	0.080	(0.049, 0.132)	0.690	(0.508, 0.939)	88
40	9	Unif	0.115	(0.065, 0.206)	0.617	(0.437, 0.871)	81
40	15	Expo	0.569	(0.342, 0.949)	1.751	(1.183, 2.591)	68
40	15	LogNorm	2.609	(1.935, 3.518)	2.563	(1.973, 3.330)	-2*
40	15	Beta	0.565	(0.378, 0.844)	0.826	(0.606, 1.126)	32*
40	15	Unif	0.831	(0.558, 1.239)	0.753	(0.487, 1.162)	-10*
200	2	Expol	0.000	(0.000, 0.001)	0.037	(0.018, 0.076)	99
200	2	LogNorm	0.001	(0.000, 0.002)	0.125	(0.064, 0.248)	99
200	2	Beta	0.000	(0.000, 0.001)	0.009	(0.004, 0.019)	99
200	2	Unif	0.000	(0.000, 0.001)	0.009	(0.004, 0.018)	99
200	9	Expo	0.010	(0.005, 0.021)	0.805	(0.526, 1.230)	99
200	9	LogNorm	0.072	(0.044, 0.118)	1.889	(1.290, 2.770)	96
200	9	Beta	0.006	(0.003, 0.012)	0.240	(0.165, 0.349)	98
200	9	Unif	0.010	(0.004, 0.021)	0.234	(0.156, 0.352)	96
200	15	Expo	0.087	(0.048, 0.159)	1.213	(0.855, 1.721)	93
200	15	LogNorm	0.460	(0.308, 0.687)	2.044	(1.560, 2.678)	77
200	15	Beta	0.064	(0.036, 0.114)	0.423	(0.315, 0.567)	85
200	15	Unif	0.103	(0.055, 0.191)	0.421	(0.283, 0.623)	76

*not statistically significant

4.4.2 Approximate Convergence

The experiment uses the approximate convergence summary metric described in Section 4.3, but a full sample path is shown for one run of the experiment in Figure 11. This sample convergence path highlights the typical behavior of both methods. The variance method tends to underestimate the true number of servers, followed by an overestimation before converging. As shown in Proposition 3 and seen in Figure 11, the order-based method never overestimates the true number of servers. In some cases, both methods perform similarly and the variance method has no overestimation. The underestimation from the order-based method is also sometimes greater than the

overestimation in the variance method. Despite these scenarios where the methods are similar, the order-based method has good overall empirical convergence properties as compared to the variance method.

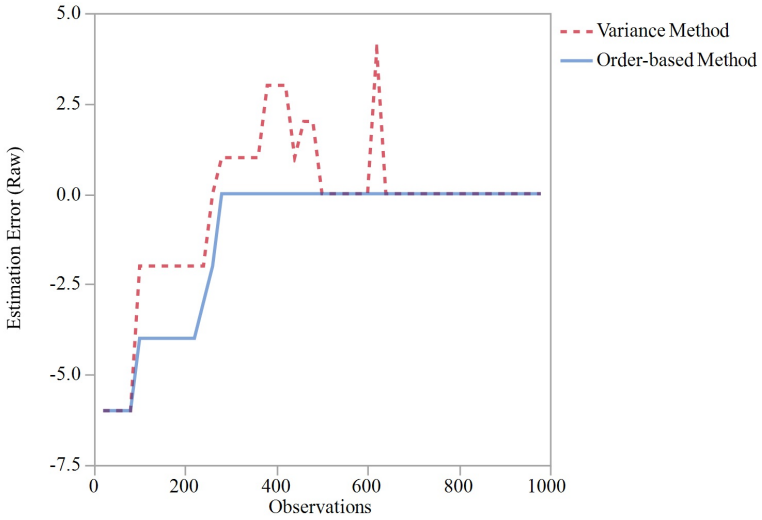

Figure 11. Sample approximate convergence path

The variance method did not reach approximate convergence within the observation limit in 203 of the 864 total runs. As a result, we have censored data for 23% of the variance method data. The order-based method reached approximate convergence for all but 12 runs. We use the censored data to produce a non-parametric simultaneous 95% confidence interval (Hall and Wellner, 1980) for the distribution of observations required to reach approximate convergence for each estimation method, shown in Figure 12. There is a statistically significant difference between estimation methods, with the order-based method requiring fewer observations to approximately converge in the majority of the design space. Over the design space considered in the experiment, the order-based method has a 50% probability of converging within 60 observations, compared to 401 observations for the variance method, an 85% reduction. In some cases, there is no distinction between the two methods. Both methods sometimes converge quickly to the correct number of servers, typically during the

runs with two servers. At the extreme upper end, the censored variance method does not provide enough data to fully estimate the upper quantiles above 0.75, while the order-based method has enough data to estimate quantiles through 0.98.

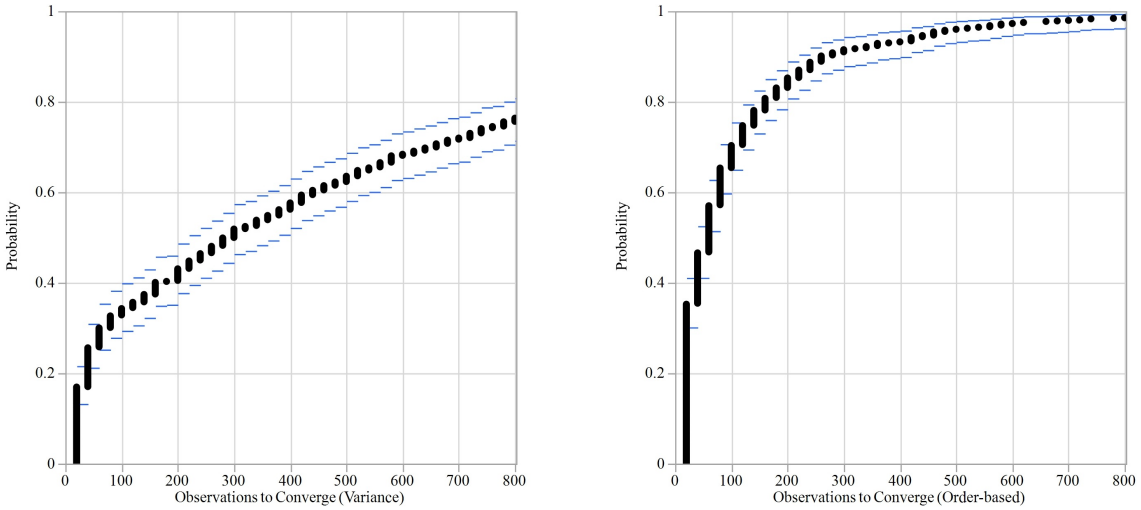

Figure 12. Approximate convergence for variance method (left) and order-based method (right)

4.4.3 Deterministic Service

In the case of deterministic service, the order-based method fails to produce meaningful estimates. This contrasts with the variance method which suffers no degradation for deterministic service queues. For the order-based method, the case with deterministic service is identifiable by identical departure and arrival order. If the data exhibit this property, the variance method is preferred.

4.4.4 Robustness

In addition to the standard case, we examine the performance of both methods when the arrival and departure times are subject to measurement error. In practice, it may be difficult to precisely capture the arrival and departure times. We introduce measurement error into the arrival and departure times by allowing the measured

times to vary uniformly between successive customers. This change introduces a high amount of time measurement error while maintaining the true order of customers. The same experiment described in Section 4.3 was conducted with the noisy data.

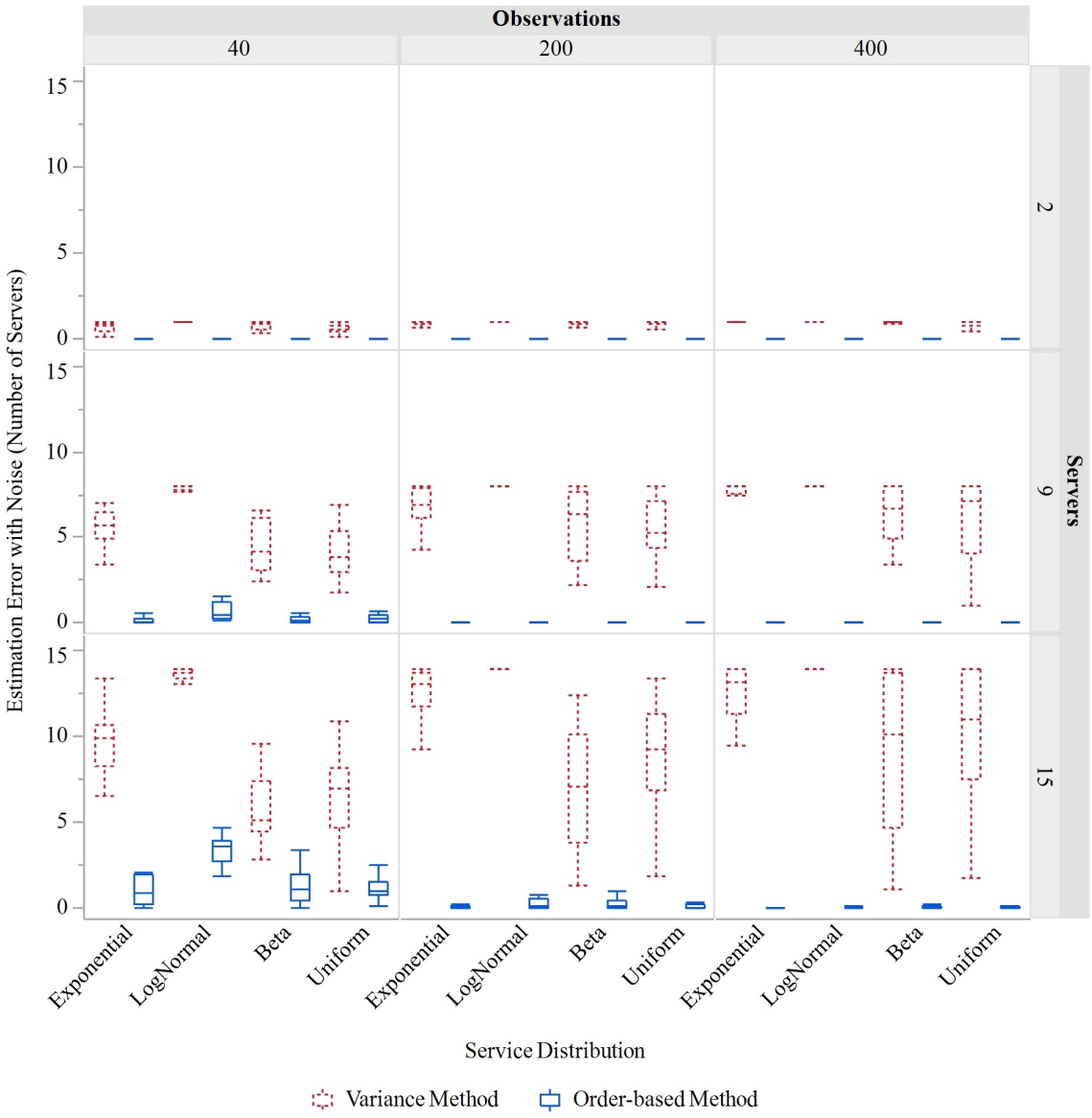

Figure 13. Average estimation error with noisy data

The order-based method is unaffected by time measurement error because it only uses arrival and departure order as inputs. The variance method, however, is highly sensitive to this amount of measurement error. In the majority of cases there is no

feasible solution to the variance minimization problem because the time error makes some estimated service times negative for any $\hat{c}_n$ value. If we relax the variance method to default to a single server if there is no feasible solution, the performance is poor. The raw results in Figure 13 suggest that the order-based method is more robust than the variance method. Overall, using the noisy arrival and departure data and a t-test, we find a statistically significant difference in the average estimation error with a p-value less than 0.001 and a 95% reduction from 5.85 servers to 0.27 servers. The same regression approach used for the non-robust case yields a a 95% confidence interval of (4.25, 5.59) for the Estimation Method[Variance] term. These results confirm the qualitative relationship seen in Figure 13. Furthermore, in 98% of runs, the variance method fails to approximately converge. In the majority of cases, the variance method has $c - 1$ estimation error because it is forced to accept the default estimate of a single server. The approximate convergence of the order-based method is not affected by the time error, as expected. These results highlight the robust performance of the order-based method with respect to arrival and departure time measurements.

We also consider the performance of both methods when the arrival and departure order are subject to measurement error. This error might be induced by the process of identifying the arriving or departing customers, for example based on the misclassification rate of facial identification or digital signatures. Unfortunately, both methods are highly sensitive to even small errors in order measurement. To represent misidentification, we randomly reassign the customer identifier for 10% of customers. This relatively minor measurement error leads to extremely poor performance for both methods. Neither method reaches approximate convergence within the observation limit for any factor settings.

Although not examined experimentally here, it is known that the variance method

performs poorly in over-served queues with low traffic intensity (Park et al., 2011). The same problem applies to the order-based method, although the order-based method does provide a lower bound, as long as customers are accurately identified. The underlying difficulty in low-traffic queues is that some servers are never or rarely busy, leading to incorrect estimators.

4.5 Last-Come, First-Served

We also extend the primary work to the non-preemptive LCFS setting. The LCFS variance minimization method developed by Park et al. (2011) is shown in Eq. (21) with additional variable definitions in Table 11. This minimization problem is that same as the minimization problem in Eq. (20), with the exception of constraint (21b).

$$\min_{\hat{c}_n} \sum_{i=1}^{n} \left(\hat{S}_i - \bar{\hat{S}}\right)^2 \tag{21a}$$

$$\text{subject to } \hat{B}_i = \begin{cases} A_i & \text{if } N_i^A < \hat{c}_n \\ D^1(N_i^A, A_i) & \text{otherwise} \end{cases}, \qquad i = 1, \ldots, n, \tag{21b}$$

$$\hat{S}_i = D_i - \hat{B}_i, \qquad i = 1, \ldots, n, \tag{21c}$$

$$\hat{S}_i > 0, \qquad i = 1, \ldots, n, \tag{21d}$$

$$1 \leq \hat{c}_n \leq n, \tag{21e}$$

$$\hat{c}_n \in \mathbb{Z}. \tag{21f}$$

The order-based algorithm does not transfer directly to the LCFS case. The LCFS discipline introduces disorder in the queue prior to service, which invalidates the framework for the order-based algorithm. However, with slight modifications

Table 11. Variable definitions (Park et al., 2011)

Variable	Definition
N_i^A	Number of customers that customer i sees upon arrival
N_i^D	Number of customers that customer i leaves behind at departure
$D^1(n,t)$	The first departure time, after time t, that leads to $N_i^D = n$

to the setting assumptions, a similar order-based idea can be extended to LCFS queues, shown in Algorithm 3 with variable and function definitions in Table 12. The *CombinedOrder* variable is a record of both arrivals and departures. It is stored as an ordered vector of customer arrival indices that includes every arrival index twice: once for the arrival and once for the departure. The vector is sorted by event order, so that the relative ordering of arrivals and departures is captured. This contrasts with the FCFS algorithm in which the order of arrivals is known and the order of departures is known, but there is no information about the relative ordering of departures with respect to arrivals or vice versa.

Algorithm 3 uses the *CombinedOrder* variable to determine exactly which customers are in the system, a lower-bound on the number of customers that must be in service, and an estimate of the number of customers in the queue (some of whom may actually be in service). The algorithm keeps a running record of which customers are in the system. When a customer leaves, the customer with the highest arrival index that is not already known to be in service either enters service or was actually already in service. In either case, we can update the estimated number of customers in service and the estimated number of servers. Table 13 shows a numerical example of the order-based LCFS algorithm.

The LCFS order-based algorithm provides a lower-bound on the number of servers and converges in probability to the correct number of servers, as stated in the following results.

Proposition 5. *Given an LCFS GI/G/c queue that meets Assumptions (i)-(vi),*

Table 12. Variable and function definitions for Algorithm 3

Name	Type	Definition
CombinedOrder	Vector of integers	Vector of arrival indices of customers ordered by increasing event times of arrivals and departures
n	Integer	Combined number of arrived and departed customers
NumServers	Integer	Estimated number of servers
CustsInSystem	Set of integers	Set of arrival indices of customers known to be in the system given the first i arrival or departure events
CustsInService	Set of integers	Set of arrival indices of customers known to be in service given the first i arrival or departure events
CustsInQueue	Set of integers	Set of arrival indices of customers estimated to be in the queue given the first i arrival or departure events
MakeSet	Function	Initialize set of integers
SetDifference	Function	Return the elements of the first set that are not in the second set
$\lvert \cdot \rvert$	Function	Return the number of elements in a set
Add	Function	Add the second argument to the set in the first argument
Maximum	Function	Maximum element in the set
Max	Function	Maximum of two integers

Algorithm 3 Order-based LCFS Server Estimation

```
procedure OrderEstimatorLCFS(CombinedOrder)
    n ← combined number of arrived and departed customers
    NumServers ← 0
    CustsInSystem ← MakeSet()
    CustsInService ← MakeSet()
    for i = 1 : n do
        if CombinedOrder[i] ∈ CustsInSystem then
            CustsInSystem ← SetDifference(CustsInSystem, CombinedOrder[i])
            CustsInService ← SetDifference(CustsInService, CombinedOrder[i])
            if |CustsInSystem| > 0 then
                CustsInQueue ← SetDifference(CustsInSystem, CustsInService)
                CustsInService ← Add(CustsInService, Maximum(CustsInQueue))
            end if
        else
            CustsInSystem ← Add(CustsInSystem, CombinedOrder[i])
        end if
        NumServers ← Max(NumServers, |CustsInService|)
    end for
return NumServers
end procedure
```

there is a positive probability that the LCFS order-based estimation algorithm produces the correct estimate immediately following the cth departure.

Proposition 6. *Given an LCFS GI/G/c queue that meets Assumptions (i)-(vi), the LCFS order-based estimation algorithm produces an estimate that is a lower bound on the true number of servers.*

Theorem 7. *For an LCFS GI/G/c queue that meets Assumptions (i)-(vi), the LCFS order-based server estimation algorithm produces estimates* $\hat{c}_n$ *such that* $\lim_{n\to\infty} P(|\hat{c}_n - c| > \epsilon) = 0 \quad \forall\, \epsilon > 0.$

The proofs for Proposition 5, Proposition 6 and Theorem 7 are included in Appendix A Section 1.1. These proofs use a similar structure to the FCFS proofs. Notably, Proposition 6 requires $2c$ arrivals for the LCFS case while Proposition 3 requires c arrivals for the FCFS case. The increase in required customers for cor-

rect estimation may be related to the different performance in the FCFS and LCFS scenarios.

Table 13. Order-based LCFS algorithm example

Customer ID	Arrival Index	Event Type	Running Estimate	Cumulative Estimate
79305	1	Arrival	1	1
89698	2	Arrival	1	1
48692	3	Arrival	1	1
79305	1	Departure	1	1
21094	4	Arrival	1	1
89698	2	Departure	2	2
71772	5	Arrival	2	2
50982	6	Arrival	2	2
71772	5	Departure	3	3

Although the LCFS case has desirable long-run properties, the experimental results suggest that the LCFS order-based estimator does not improve upon the LCFS variance estimator. The LCFS algorithms were compared using the same experimental design as in the FCFS case, shown in Table 8. Although four total replications of the design were conducted for the FCFS case, two total replications were conducted for the LCFS case. Two replications were sufficient to show the similar performance of the LCFS order-based method and the LCFS variance method. Both estimators require a similar number of observations to reach approximate convergence, with no statistically significant difference at the 95% confidence level, as seen in Figure 15. The estimation error of the LCFS order-based method is generally similar to the error for the LCFS variance method, but in some cases there is a statistically significant degradation from the LCFS variance method to the LCFS order-based method. Figure 14 shows the empirical differences in estimation error. Statistical regression models confirm the empirical trend that the LCFS-order based method does not improve on the LCFS variance method. This lack of improvement might be related to the increased disorder in the LCFS input queue as compared to the FCFS input

queue. Since the order-based estimator is conservative, the estimate increases by at most a single server after a departure in the LCFS setting. In the FCFS setting, however, the estimate can sometimes jump by several servers after a single departure by taking advantage of the stable input order. Since the variance of service times is unaffected by queue discipline, the variance method does not share the same sensitivity to LCFS queues.

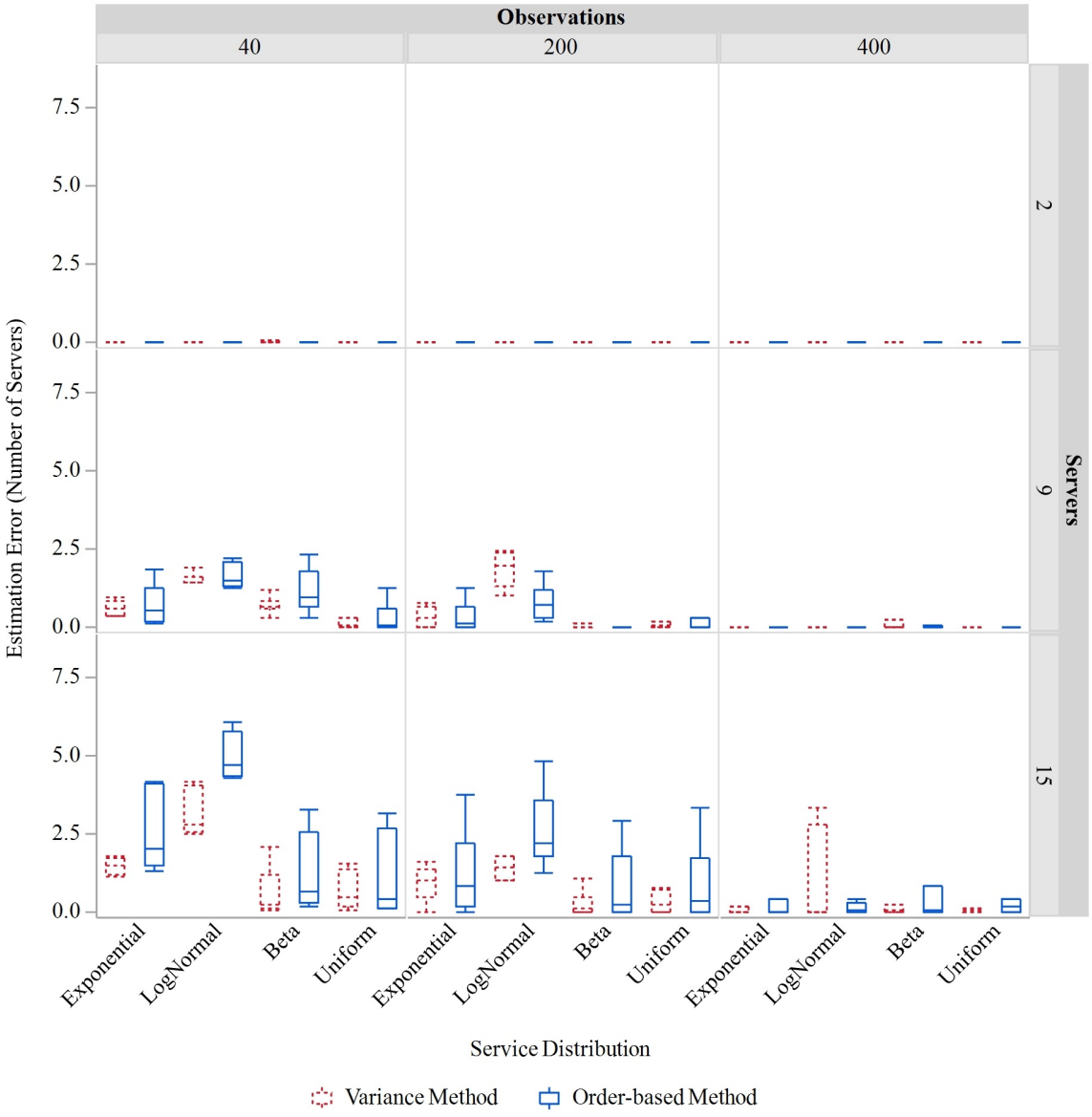

Figure 14. Average LCFS estimation error

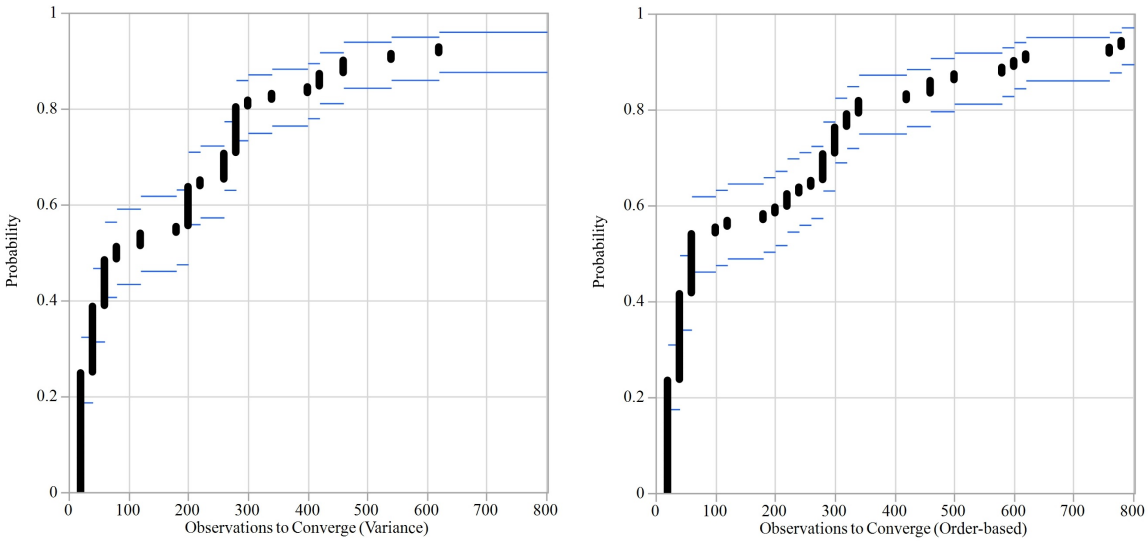

Figure 15. Approximate convergence for LCFS variance method (left) and LCFS order-based method (right)

Another major difference between the FCFS order-based method and the LCFS order-based method is that the LCFS order-based method is less robust than the FCFS order-based method. Any time errors that do not affect the combined arrival and departure order have no effect on the LCFS order-based estimate. However, the LCFS order-based estimator depends on the combined arrival and departure order, while the FCFS order-based estimator depends on the arrival order and departure order separately. Compared to the combined order, the separate orders allow for more time measurement error without affecting the overall order. This distinction limits the amount of error that the LCFS version can address. Furthermore, the LCFS variance method is also robust to these types of small errors in the departure time, as seen in Figure 16. In the LCFS context, both methods are able to achieve good results with small departure error that does not affect the combined event order. Overall, the performance of the LCFS order-based estimator does not improve upon the performance of the LCFS variance estimator.

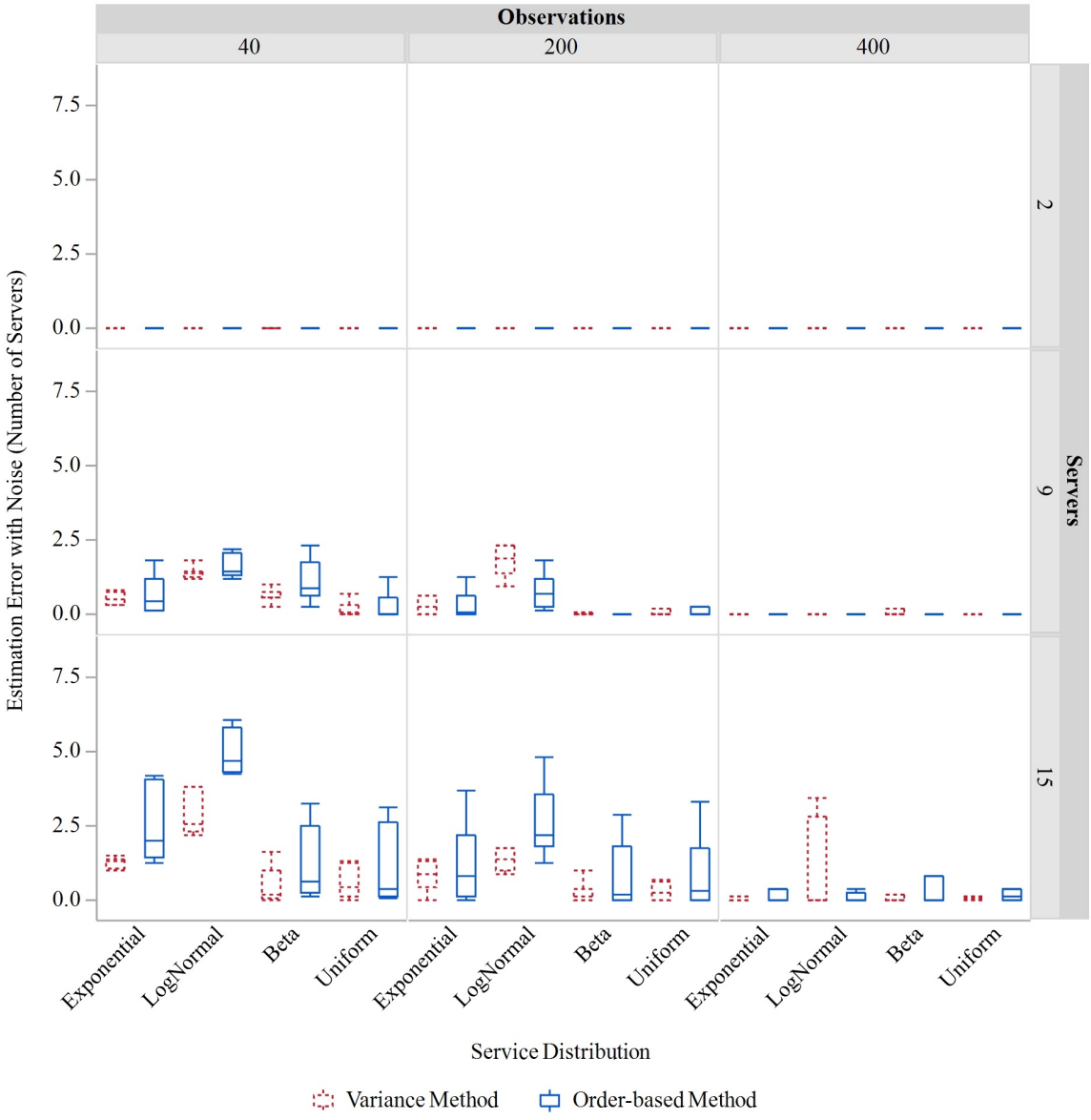

Figure 16. Average LCFS estimation error with noisy data

4.6 Conclusion

In this chapter, we propose a robust approach for queue inference when the internal parameters of an FCFS GI/G/c queue are unknown and the service is non-deterministic. Compared to the existing variance minimization method, our experimental results for a relevant design space show that, in the FCFS setting, the order-based method reaches approximate convergence more quickly and has similar or lower error before converging. Theoretical results show that the order-based method has desirable long-run properties in general for the setting of interest. Furthermore, the order-based method is robust to measurement error in arrival and departure times. These characteristics make the order-based method a suitable estimator for small sample sizes with noisy measurements and limited knowledge of the internal structure of the queue. Although the theoretical results can be extended to the LCFS setting, the LCFS order-based estimator does not improve on the previous methods with respect to empirical performance for small, noisy samples.

The results presented here could be extended in several ways. A preliminary exploration of entropy minimization as an alternative to variance minimization suggests that there is no improvement over the variance minimization approach, but there may be implementation choices that lead to better results. The current methods could also be enhanced with measures of statistical confidence for each sample size or expected time to converge to a given precision. New methods to solve this problem could also provide value by generalizing the context. Approaches that are able to address partially observable or unobservable customer identity would allow for application to a broader class of problems. It would also be desirable to relax the setting restrictions by extending these results to consider queueing networks rather than simple queues.

For GI/G/c queues with unobservable service in the LCFS setting, the order-based method and variance method have similar performance but the order-based

method appears to display larger error in some cases. In the FCFS setting, the order-based method is a robust estimator for the number of servers in a GI/G/c queue with unobservable service. This estimator has good empirical performance for simulated small, noisy samples and has desirable theoretical convergence and lower-bound properties.

V. Robust Sequential Optimization Under Uncertainty

5.1 Introduction

Markov decision processes (MDPs) have been widely used for complex planning problems under uncertainty. However, in many practical applications, the implementation environment differs from the planning model due to noise or incomplete information, particularly with respect to the model dynamics. The decision maker may be unable to directly observe the state of the environment and may be uncertain about the precise transition, observation or reward functions. These challenges introduce a second layer of ambiguity in addition to the traditional probabilistic uncertainty in MDPs. The ideal policy in this setting should be robust to uncertainty in any of the model components.

Modeling incomplete information is particularly relevant for sequential decision making problems that have a goal of collecting information. In MDP and partially observable MDP (POMDP) formulations with utility-based rewards, environmental uncertainty results in information-seeking strategies as an intermediate step to maximizing rewards. When the rewards are purely information-based, however, information-seeking strategies are the primary purpose rather than an intermediate step. Applications with information-based rewards include active sensing, situational awareness, surveillance, and detection problems. Modeling second-order knowledge about the problem dynamics is critical in these information-based applications.

This chapter directly addresses the degraded performance of POMDPs under model misspecification by developing robust belief-reward partially observable Markov decision processes. Our main contributions include the following. We present a new formulation for robust belief-reward POMDPs. We present a robust belief-reward point-based value iteration algorithm for addressing this general class of POMDPs.

We include theoretical results which show that our algorithm has desirable convexity and convergence properties. We conduct experiments to empirically evaluate the sensitivity of traditional POMDPs to model misspecification and the improved performance of robust belief-reward POMDPs. Lastly, we apply the robust belief-reward formulation and solution technique to a practical cybersecurity resource allocation problem and show that the robust algorithm outperforms the standard algorithm in the presence of model uncertainty.

In Section 5.2, we provide background on existing modeling approaches. Then, in Section 5.3, we introduce the formulation for a new model, robust belief-reward POMDPs, and introduce the robust belief-reward point-based value iteration algorithm. In Section 5.4, we present an experiment comparing the standard, robust, belief-reward, and robust belief-reward formulations. In Section 5.5, we apply the robust belief-reward approach to a cybersecurity intrusion detection problem with realistic limitations on model precision. Lastly, in Section 5.6, we summarize the findings and identify areas for future research.

5.2 Partially Observable Markov Decision Processes

This section provides background on two partially-observable generalizations of MDPs. POMDPs extend standard MDPs by incorporating state uncertainty through an observation model (White, 1991; Cassandra et al., 1994). Belief-reward POMDPs generalize standard POMDPs by allowing for reward functions that depend on the true states and actions as well as on the decision maker's beliefs. Robust POMDPs generalize standard POMDPs by allowing ambiguity in the transition and observation functions.

5.2.1 Belief-Reward and Information-Reward POMDPs

In some domains, the reward is a function of the state uncertainty rather than the state itself. In principle, this situation can be represented with standard POMDPs, where future rewards drive information gathering in the short term. This approach has been successful for military sensing problems (Yost and Washburn, 2000). However, when the standard actions and information-gathering actions are sufficiently decoupled, this formulation is not attractive. For instance, in the active sensing problem area, rewards are most naturally modeled as a function of the information change (Mihaylova et al., 2003). Furthermore, in surveillance applications, the information collection problem is well-defined, but the available actions and non-information rewards based on the collected information are subject to an ill-structured decision process involving collective human judgment. Traditional POMDPs allow sensing costs and task-dependent rewards, but in these information-focused applications, belief-dependent rewards are more appropriate.

Information-reward and belief-reward POMDPs modify standard POMDPs to allow non-standard rewards. Although closely related, belief-reward POMDPs and information-reward POMDPs take different modeling approaches. The belief-reward approach introduces a modified reward function while the information-reward approach modifies the state space definition of standard POMDPs.

Araya-López et al. (2010) introduce belief-reward POMDPs, also called ρPOMDPs. Araya-López et al. (2010) define the ρPOMDP with respect to the generalized belief-dependent reward function, $\rho(\mathbf{b}, a)$ rather than the standard state-dependent reward function $r(s, a)$, where s is a state, a is an action, and b is a belief vector over states. Convex belief-dependent reward functions maintain convexity of the value function, as shown by Araya-López et al. (2010, Lemma A.1) and Araya-López et al. (2010, Theorem 3.1). Furthermore, piecewise linear convex (PWLC) belief-dependent reward

functions maintain a PWLC value function for the overall ρPOMDP (Araya-López et al., 2010). This property is exploited by optimal solution techniques, such as incremental pruning (Cassandra et al., 1997; Zhang and Liu, 1996), and by point-based approximate solution techniques (Lovejoy, 1991; Pineau et al., 2003; Smith and Simmons, 2005; Kurniawati et al., 2008). Convex belief-dependent reward functions violate the PWLC assumption of most POMDP solvers, but they can be approximated by PWLC functions and the error introduced by the approximation is bounded (Araya-López et al., 2010). These results allow for information-measure reward functions such as approximate negative information entropy or approximate Kullback-Leibler divergence.

Spaan et al. (2015) introduce POMDPs with information rewards using a different modeling approach than belief-reward POMDPs. POMDPs with information rewards are structured to meet the criteria of a standard POMDP while still allowing information-dependent rewards. This result is accomplished by augmenting the action space with additional *commitment* actions that allow the decision maker to take actions that represent assertions about the value of important states (Spaan et al., 2015).

Eck and Soh (2012) also review several approaches to hybrid rewards that include a belief-dependent component and a state-action component. The traditional state-action component is often a negative reward associated with the cost of sensing. For example, using the negative expected entropy, the hybrid reward, R_h, is a weighted sum of the state-action component, $R_s = \sum_{s \in \mathcal{S}} \mathbf{b}(s) R(s,a)$, and the belief-based component, $R_{\mathbf{b}} = \mathbb{E}\left[H(\mathbf{b}_{a,z})\right]$,

$$R_h(\mathbf{b}, s, a, z) = wR_s + (1-w)R_{\mathbf{b}} \tag{22}$$

$$= w \sum_{s \in \mathcal{S}} \mathbf{b}(s) r(s,a) - (1-w)\mathbb{E}\left[H(\mathbf{b}_{a,z})\right], \tag{23}$$

where w is the weight, $r(s, a)$ is the state-reward, and $H(\mathbf{b}_{a,z})$ is the entropy given action a and observation z (Eck and Soh, 2012). In the remainder of this chapter, we focus on the belief-reward approach to avoid the penalty of augmenting the state space. Note that belief-reward POMDPs are a strict generalization of POMDPs.

5.2.2 Robust POMDPs

Robust MDPs extend the traditional MDP framework for sequential decision making under uncertainty to sequential decision making under *ambiguity*, where ambiguity is probabilistic uncertainty with an unknown distribution (Camerer and Weber, 1992). Early work on MDPs with ambiguous dynamics was conducted by Satia and Lave (1973), White and Eldeib (1986), and White and Eldeib (1994). More recent work addresses MDPs using a formal robust framework (Nilim and El Ghaoui, 2005; Iyengar, 2005).

In the partially observable setting, robust POMDPs extend POMDPs by allowing ambiguity in the specification of the transition and observation probabilities (Osogami, 2015). This ambiguity may be caused by noisy data, incomplete knowledge, or conflicting input from subject matter experts. Robust POMDPs combine principles from both robust MDPs (e.g., Kaufman and Schaefer (2013)) and POMDPs (e.g., Kurniawati et al. (2008)). Specifically, the known transition and observation probabilities are replaced with ambiguous transition and observation probabilities that are only known to belong to an ambiguity set. The robust solution is optimal in the worst case scenario, i.e., under the least favorable probabilities in the ambiguity set. While the robust solution provides protection against worse-than-expected model dynamics, the performance in the expected case is generally worse than the non-robust solution. The degree of conservatism can be controlled by the size of the ambiguity set, with larger ambiguity sets producing more conservative solutions. Note that ro-

bust POMDPs generalize POMDPs: when the ambiguity set is a degenerate set of only a single transition-observation function, we recover a standard POMDP.

5.3 Robust Belief-Reward Partially Observable Markov Decision Processes

In this section, we introduce the formulation for robust belief-reward POMDPs. Robust belief-reward POMDPs generalize the reward, dynamics, and observability of standard MDPs. For a robust belief-reward POMDP in which the ambiguity set for the dynamics is degenerate, the belief-reward is degenerate, and the observation function is perfect, a standard MDP is recovered. However, in many practical applications the model dynamics are not known precisely and MDP and POMDP solutions are sensitive to misspecification (Nilim and El Ghaoui, 2005; Itoh and Nakamura, 2007). The robust belief-reward POMDP formulation addresses this shortcoming by accounting for model misspecification in information acquisition problems.

5.3.1 Formulation

The robust belief-reward POMDP, or robust ρPOMDP, is defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{Z}, \mathcal{P}, \rho, \gamma)$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{Z}$ is the observation space, $\mathcal{P}$ is the ambiguity set of transition-observation functions, ρ is the belief-reward function, and γ is the discount factor. We consider POMDPs with discrete state, action, and observation spaces over discrete time.

The ambiguity set, $\mathcal{P}$, includes all $p_n^a(s', z|s)$ in the given ambiguity set, where $p_n^a(s', z|s)$ is defined as the probability of transitioning to $s' \in \mathcal{S}$ and observing $z \in \mathcal{Z}$ given that $a \in \mathcal{A}$ is taken from $s \in \mathcal{S}$ at time $n \in [0, N]$. For a given state, s, and action, a, the ambiguity set $\mathcal{P}_s^a$ is a set of categorical probability distributions of dimension $|\mathcal{S}| \times |\mathcal{Z}|$. These ambiguity sets can take a variety of forms (e.g., interval,

polytope, ellipsoid, norm-based) but are restricted to convex sets for value iteration solution techniques.

The belief-dependent reward, ρ, is a function of the action, a and the belief-state vector, $\mathbf{b} \in \mathcal{B}$, where $\mathcal{B}$ is the continuous belief space defined over the discrete state space, $\mathcal{S}$. The updated belief is represented by $\mathbf{b}'_{\mathbf{b},a,z}$ where $\mathbf{b}$ is the previous belief, a is the action selected, and z is the observation received. The discount factor, γ, is a constant parameter that controls the time-value of rewards.

Given this model, we can construct the value function from the belief-reward value function and the robust value function. The value function for the belief-reward POMDP is

$$V_n(\mathbf{b}) = \max_{a \in \mathcal{A}} \rho(\mathbf{b}, a) + \sum_{s \in \mathcal{S}} \left(\mathbf{b}(s) \gamma \sum_{t,z \in \mathcal{S} \times \mathcal{Z}} p_n^a(t, z|s) V_{n+1}(\mathbf{b}'_{\mathbf{b},a,z}) \right), \tag{24}$$

the value function for the robust POMDP is

$$V_n(\mathbf{b}) = \max_{a \in \mathcal{A}} \min_{p_n^a \in \mathcal{P}^a} \sum_{s \in \mathcal{S}} \mathbf{b}(s) \left(r(s, a) + \gamma \sum_{t,z \in \mathcal{S} \times \mathcal{Z}} p_n^a(t, z|s) V_{n+1}(\mathbf{b}'_{\mathbf{b},a,z}) \right), \tag{25}$$

and the value function for the robust belief-reward POMDP can be directly derived from both of these value functions as

$$V_n(\mathbf{b}) = \max_{a \in \mathcal{A}} \min_{p_n^a \in \mathcal{P}^a} \rho(\mathbf{b}, a) + \sum_{s \in \mathcal{S}} \left(\mathbf{b}(s) \gamma \sum_{t,z \in \mathcal{S} \times \mathcal{Z}} p_n^a(t, z|s) V_{n+1}(\mathbf{b}'_{\mathbf{b},a,z}) \right). \tag{26}$$

This chapter introduces two results that provide a foundation for solving robust belief-reward POMDPs. First, using the result proven by Osogami (2015, Theorem 1), the robust belief-reward value function is convex. The proof for Theorem 8 is included in Appendix B.

Theorem 8. *When $N < \infty$, the robust belief-reward value function, $V_n(\mathbf{b})$, is convex*

with respect to $\mathbf{b}$ *for each* $n \in [0, N]$ *for a convex ambiguity set,* $\mathcal{P}_s^a$ *for* $s, a \in \mathcal{S} \times \mathcal{A}$*, and a convex belief-reward function,* $\rho(\mathbf{b}, a)$*.*

Second, using the result proven by Osogami (2015, Theorem 2), the value function converges. The proof for Theorem 9 is included in Appendix B.

Theorem 9. *The robust belief-reward value function,* V_0*, satisfying*

$$V_n(\mathbf{b}) = \max_{a \in \mathcal{A}} \rho(\mathbf{b}, a) + \min_{p_n^a \in \mathcal{P}^a} \sum_{s \in \mathcal{S}} \left(\mathbf{b}(s) \gamma \sum_{t,z \in \mathcal{S} \times \mathcal{Z}} p_n^a(t, z|s) V_{n+1}(\mathbf{b'}_{\mathbf{b},a,z}) \right), \tag{27}$$

converges uniformly as $N \to \infty$ *if* $\gamma < 1$*.*

Theorem 8 and Theorem 9 allow classic value iteration techniques to be applied to the robust belief-reward setting, while introducing only mild conditions on the model.

5.3.2 Approximate Value Iteration for Robust Belief-Reward POMDPs

In this section, we introduce an approximate value iteration algorithm for robust belief-reward POMDPs. Using the results proven by Araya-López et al. (2010) and Osogami (2015), we show that value iteration is valid for robust belief-reward POMDPs. Value iteration is particularly attractive in the standard POMDP setting because the reward function is PWLC, which allows for ϵ-optimal solutions to be found in finite time despite the infinite dimension of the belief space in the underlying MDP. While continuous space MDPs are generally intractable, the special structure of the belief space and the original POMDP supports efficient state transitions, or *belief updates*, and guarantees a piecewise linear and convex (PWLC) value function (Kochenderfer, 2015). The structural properties allow for ϵ-optimal solution techniques for small-sized POMDP problems (White and Eldeib, 1994; Zhang and Liu, 1996; Cassandra et al., 1997).

In the belief-reward POMDP setting, if the belief-dependent reward is convex, the value function is convex; if the belief-dependent reward is PWLC, the value function is PWLC (Araya-López et al., 2010). As a result, ϵ-optimal solutions are available for PWLC belief-reward POMDPs. However, in the robust POMDP setting, if the ambiguity set for the dynamics, $\mathcal{P}$, is convex, the value function is convex (Osogami, 2015). This requirement restricts robust POMDP value iteration to approximate solutions using a PWLC approximation of the value function.

These results imply two restrictions for the robust belief-dependent case: the belief dependent reward, $\rho(\mathbf{b}, a)$, must be PWLC, and the ambiguity set, $\mathcal{P}$, must be a convex set. Convexity is a mild restriction on the ambiguity set, which allows upper and lower probabilities, p-boxes, and polytopic uncertainty, among others.

In practice, the restriction to approximate value iteration is typically not a significant limitation because computational complexity limits the problem size for exact value iteration. Under the restriction to PWLC information-reward functions, $\rho(\mathbf{b}, a)$ can be represented by a set of vectors for each action, Γ^a, such that

$$\rho(\mathbf{b}, a) = \max_{\alpha_\rho \in \Gamma^a} \sum_{s \in \mathcal{S}} \mathbf{b}(s)\alpha_\rho(s). \tag{28}$$

This representation is analogous to the alpha-vector representation of POMDP value functions (Araya-López, 2013). Point-based value iteration algorithms update the value function over a finite subset, $\mathcal{B}_0$, of the full belief space $\mathcal{B}$. However, each point-based update corresponds to an alpha vector over the entire belief space, so the policy is fully defined over the original belief space.

To address the robust requirement, we use the argument developed by Osogami (2015), where equation (29) induces a min-max game against nature in which nature selects the probability distribution that minimizes rewards for the decision maker. The estimated value function at iteration $n+1$, $\hat{\Lambda}_{n+1}$, is represented as a set of alpha

vectors, resulting in the following definition for the value-minimizing probability at iteration n,

$$p_n^{a,*} = \underset{p_n^a \in \mathcal{P}^a}{\operatorname{argmin}} \sum_{z \in \mathcal{Z}} \max_{\alpha_z \in \hat{\Lambda}_{n+1}} \sum_{s \in \mathcal{S}} \mathbf{b}(s) \sum_{t \in \mathcal{S}} p_n^a(t, z|s) \alpha_z(t). \tag{29}$$

This game results in the minimization problem shown in equation (30), where $U(z)$ acts as a bounding variable in the optimization formulation,

$$\begin{aligned}
\underset{p_n^a(\cdot,\cdot|s)}{\text{minimize}} \quad & \sum_{z \in \mathcal{Z}} U(z) && \text{(30a)} \\
\text{subject to} \quad & U(z) \geq \sum_{s \in \mathcal{S}} \mathbf{b}(s) \sum_{t \in \mathcal{S}} p_n^a(t, z|s) \alpha_z(t), \quad \forall \alpha_z \in \hat{\Lambda}_{n+1}, \forall z \in \mathcal{Z}, && \text{(30b)} \\
& p_n^a(\cdot,\cdot|s) \in \mathcal{P}_s^a, \quad \forall s \in \mathcal{S}. && \text{(30c)}
\end{aligned}$$

This minimization problem is a linear program for appropriate choices of the ambiguity sets, $\mathcal{P}_s^a$, for instance interval or polytopic ambiguity. The maximization in equation (29) selects the α_z^* determined by equation (31) for the solution to equation (30),

$$U(z) = \sum_{s \in \mathcal{S}} \mathbf{b}(s) \sum_{t \in \mathcal{S}} p_n^a(t, z|s) \alpha_z^*(t). \tag{31}$$

Building from these previous results, we adapt the algorithms examined by Araya-López et al. (2010) and Osogami (2015) to construct Algorithms 4 and 5 for robust belief-reward point-based value iteration. Algorithm 4 is the overall robust belief-reward point-based value iteration component. Algorithm 5 is the robust belief-reward dynamic programming component. The key change required to extend robust dynamic programming to the belief-reward case is to replace the standard reward with belief-reward by adding Step 12 based on the alpha vector definition of the belief-reward in equation (28) and by replacing the reward function in Step 13. Given that these rewards are modified, the alpha vectors in the approximate value function at each iteration reflect the value over belief-rewards rather than standard rewards.

Note that the alpha vector definition of the reward function includes both standard state-rewards and hybrid rewards as special cases, where the hybrid costs are modeled through the Γ^a sets such that $\Gamma^{a_1} \neq \Gamma^{a_2}$ and so on.

Algorithm 4 Robust belief-reward point-based value iteration

input $\mathcal{B}_0$
$\hat{\Lambda}_N \leftarrow \{\mathbf{0}\}$
for $n \leftarrow N-1:0$ **do**
    $\hat{\Lambda}_n \leftarrow$ Robust belief-reward point-based backup value of $\hat{\Lambda}_{n+1}$
end for
return $\hat{\Lambda}_0$

Algorithm 5 Robust belief-reward point-based dynamic programming backup

input $\hat{\Lambda}_{n+1}, \mathcal{B}_0$
$\hat{\Lambda}_n \leftarrow \emptyset$
for all $\mathbf{b} \in \mathcal{B}_0$ **do**
    $\hat{\Lambda}_{n,\mathbf{b}} \leftarrow \emptyset$
    for all $a \in \mathcal{A}$ **do**
        Solve the robust min-max problem (30) for $\mathbf{b}, a$
        for all $z \in \mathcal{Z}$ **do**
            $\alpha_z^* \leftarrow \alpha_z$ in the optimal solution of (30) that achieves (31)
        end for
        for all $s \in \mathcal{S}$ **do**
            $p_n^{a,*}(\cdot,\cdot|s) \leftarrow$ minimizer $p^a(\cdot,\cdot|s)$ in the optimal solution of (30)
            $\alpha_\rho \leftarrow \text{argmax}_{\alpha \in \Gamma^a} \mathbf{b} \cdot a$
            $\alpha^*(s) \leftarrow \alpha_\rho(s) + \gamma \sum_{t \in \mathcal{S}} \sum_{z \in \mathcal{Z}} p^{a,*}(t,z|s) \alpha_z^*(t)$
        end for
        $\hat{\Lambda}_{n,\mathbf{b}} \leftarrow \hat{\Lambda}_{n,\mathbf{b}} \cup \{\alpha^*\}$
    end for
    $\hat{\Lambda}_n \leftarrow \hat{\Lambda}_n \cup \left\{ \text{argmax}_{\alpha \in \hat{\Lambda}_{n,\mathbf{b}}} \sum_{s \in \mathcal{S}} \mathbf{b}(s) \alpha(s) \right\}$
end for
return $\hat{\Lambda}_n$

The standard point-based dynamic programming backup algorithm, at iteration n, requires $O(|\mathcal{B}||\mathcal{A}||\mathcal{Z}||\mathcal{S}||\hat{\Lambda}_{n+1}|)$ operations (Pineau et al., 2003). The robust belief-reward version, Algorithm 5, can also be solved in polynomial time at iteration n, but with an added linear program in the inner-most loop of the algorithm. Assuming

an interval ambiguity set, the robust calculations, shown in Algorithm 5 Steps 5 - 14, replace $O(|\mathcal{Z}||\mathcal{S}|)$ operations in the standard algorithm with three polynomial time computations: a linear program with $|\mathcal{Z}||\mathcal{S}|^2+|\mathcal{Z}|$ decision variables and $|\mathcal{Z}|+|\mathcal{S}|+2$ constraints, a loop with $O(|\mathcal{Z}||\hat{\Lambda}_{n+1}|)$ operations, and a loop with $O(|\mathcal{S}|(|\Gamma|+|\mathcal{Z}||\mathcal{S}|))$ operations. Linear programs can be solved to optimality in polynomial time, so the overall robust algorithm can be calculated in polynomial time, but there is still a substantial robustness penalty with respect to time-complexity.

5.4 Experiments

This section addresses makes two contributions. First, while robust POMDPs have been proposed in the literature, there is a lack of empirical evidence that demonstrates the performance of robust POMDPs as compared to standard POMDPs. This section uses experiments to address the gap in the literature for robust POMDPs. Second, we conduct another set of experiments for robust belief-reward POMDPs to demonstrate their improvement in the generalized belief-reward setting. Osogami (2015) includes testing on a single instance of the Heaven and Hell problem but does not systematically compare the robust POMDP performance to standard POMDP performance. Itoh and Nakamura (2007) conduct comprehensive baseline experiments for POMDPs with imprecise parameters (POMDPIPs), but POMDPIPs have substantive differences from robust POMDPs. The solution technique used in Itoh and Nakamura (2007) selects any admissible probability distribution from the ambiguity set, rather than taking a max-min utility approach from the robust literature. The primary goal of their admissible approach is to reduce computational complexity while selecting distributions from a set of acceptable distributions, rather than guaranteeing performance.

This computational experiment uses a set of standard problems from the POMDP

literature which we extend to the robust and belief-reward settings. First, we calculate the expected total discounted reward for each problem in the standard POMDP setting using point-based value iteration (PBVI). Then, we use robust point-based value iteration (RPBVI) to calculate the pessimistic expected total discounted reward subject to ambiguity in the transition and observation functions. We compare the standard and robust policies by calculating the mean simulated value of each policy. The standard policy is optimal in the nominal setting, so, by construction, it outperforms the robust policy in the nominal setting. In the ambiguous setting, the robust policy provides protection against the worst-case, while the nominal policy is degraded because of model misspecification. Additionally, we investigate the relative performance of the nominal and robust policies as the ambiguity in the problem dynamics increases.

We consider ambiguity sets for the transition and observation probability distributions defined as upper and lower probabilities, or probability intervals. We saturate the ambiguity set at ϵ and $1 - \epsilon$ where $\epsilon = 1 \times 10^{-6}$. Fully saturated bounds are disallowed to enable a comparison of the performance of the robust solution and the nominal solution when the dynamics follow the nominal case. The experiment also includes two different belief-dependent rewards. The simple belief-dependent reward is defined as a linear cone such that the extremes of the belief simplex have value 1 and the center of the belief simplex has value 0. The complex belief-dependent reward is a PWLC function with variable gradient.

To conduct the experiments, we extend the JuliaPOMDP suite of POMDP models, solvers, and tools (Egorov et al., 2017; Lubin and Dunning, 2015). Specifically, we implement a new formulation for robust POMDPs and robust belief-reward POMDPs; implement a new solver for robust point-based value iteration and robust belief-reward point-based value iteration; formulate robust, belief-reward, and robust belief-reward

models for several problems; and extend the tools necessary for interfaces, simulation, and history recording to the robust and belief-dependent settings. The algorithm implementation, model formulations, and extended tools are available at `https://github.com/ajkeith/RobustValueIteration`.

5.4.1 Robust POMDP Experiments

This section presents results comparing robust POMDPs to standard POMDPs. We consider several variants of two reference problems from the literature: the Tiger Problem (Cassandra et al., 1994) and the Crying Baby Problem (Kochenderfer, 2015). These problems are selected to facilitate comparison with other approaches due to the manageable size and the ability to interpret policy solutions. Details on the problems are available in Appendix B.

To demonstrate correctness, we compare the results of our algorithm to the results of the SARSOP algorithm for the Crying Baby Problem (Kurniawati et al., 2008). The optimal policy for the Crying Baby Problem is to do nothing until the belief about the baby's hunger reaches a certain threshold, then feed the baby. For the default parameters, the threshold is approximately $p(\text{State} = \text{Hungry}) = 0.28$. Figure 17 shows the value function for both the nominal SARSOP policy and the robust PBVI policy under several levels of increasing ambiguity. The RPBVI functions in the legend of Figure 17 are listed in decreasing order of average value. The overall value function decreases as the ambiguity set increases. The transition point in belief-space from "do nothing" to "feed" also shifts toward 0.5.

5.4.1.1 Parameters

For each decision problem, we vary the solution technique, the ambiguity level of the modeled problem (solution), and the ambiguity level of the true environment

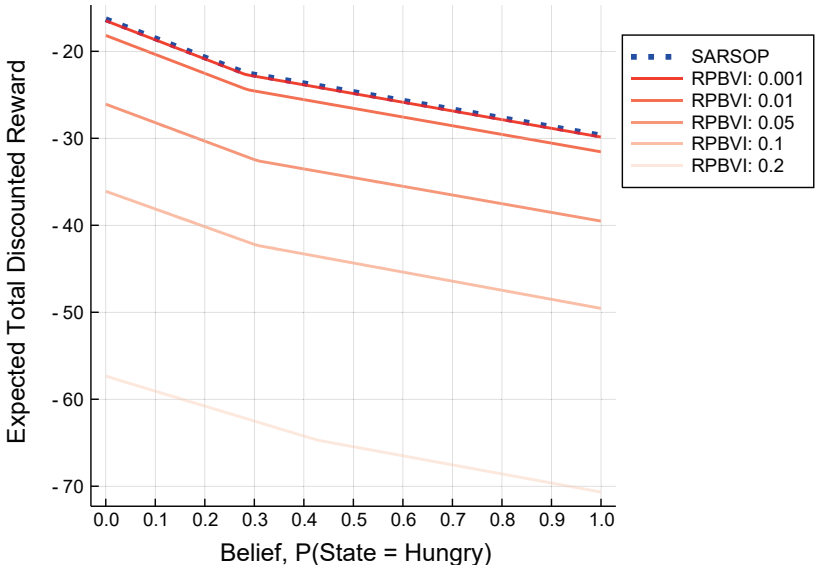

Figure 17. Robust baby POMDP value function

(simulation). These combinations correspond to columns 1 through 4 in Table 14. The fifth and sixth columns report the mean and 95% confidence interval for the simulated total discounted reward produced by each policy. The seventh and eighth columns report the accuracy of the policy. For any belief vector, the most-likely estimator of the true state is the state with the highest belief in the belief vector. By calculating this estimate at every simulation step, we can compute the percent of time that each policy's belief-state correctly predicts the true state. We refer to this metric as the accuracy metric.

In the standard version of both problems, there is no ambiguity. For the ambiguous versions, we take the original problem's dynamics as the nominal dynamics and construct an ambiguity set around the nominal transition-observation distribution. We vary the size of this ambiguity set from 0.001 to 0.2. The full details of the problem parameters are included in Appendix B.

5.4.1.2 Results

The baseline results are provided by the performance of the standard solution on the nominal model, where both the model ambiguity and the environment ambiguity are set to zero, and by the performance of the robust solution on the ambiguous model, where the model ambiguity and the environment ambiguity are non-zero. We collect these baseline results for each level of ambiguity. Then, we consider each solution technique under model misspecification. That is, we consider the performance of the nominal solution computed under the assumption of nominal model dynamics when the true environment exhibits off-nominal, worst-case ambiguity in the transitions and observations. We also consider the opposite perspective for the robust solution. That is, we consider the performance of the robust solution computed under the assumption of worst-case ambiguity for the model dynamics when the true environment exhibits nominal transitions and observations.

For each variant of the problems, we compute an approximate policy using 200 solution iterations and 63 uniformly spaced belief points. Then, we simulate the performance of the policy for the nominal setting and for the worst-case setting. For the simulated value, we conduct 500 replications of 100 decision epochs. The transitions and observations are generated from the worst-case distributions in the ambiguity set or from the nominal distribution in the ambiguity set, depending on the simulation ambiguity. Table 14 displays the results of these tests.

There are several notable trends in Table 14. For the Tiger Problem, the robust policy has good performance under the worst-case dynamics as compared to the standard policy. For an ambiguity parameter of 0.1, the robust policy has a 101% relative improvement in mean simulated value for the worst-case dynamics. For an ambiguity parameter of 0.2, the robust policy has an 83% improvement in mean simulated value for the worst-case dynamics. However, the price of robustness is

Table 14. Robust POMDP and standard POMDP experimental results (N = 500)

Problem	Solution	Ambiguity (Solution)	Ambiguity (Simulation)	Sim Value (Mean)	Sim Value (95% CI)	Sim Accuracy (Mean)	Sim Accuracy (95% CI)
Baby	Standard	0.0	0.0	-18.357	(-19.297, -17.418)	87.4	(87.1, 87.7)
Baby	Standard	0.0	0.001	-22.341	(-23.427, -21.256)	81.8	(81.4, 82.1)
Baby	Standard	0.0	0.01	-25.169	(-26.313, -24.026)	80.9	(80.5, 81.2)
Baby	Standard	0.0	0.1	-46.541	(-47.757, -45.325)	64.8	(64.3, 65.3)
Baby	Standard	0.0	0.2	-68.975	(-70.153, -67.796)	45.9	(45.4, 46.4)
Baby	Robust	0.001	0.0	-16.76	(-17.611, -15.908)	87.6	(87.3, 87.9)
Baby	Robust	0.01	0.0	-18.849	(-19.754, -17.945)	87.5	(87.2, 87.8)
Baby	Robust	0.1	0.0	-21.232	(-21.977, -20.486)	87.8	(87.5, 88.1)
Baby	Robust	0.2	0.0	-25.152	(-25.704, -24.600)	63.3	(63.1, 63.6)
Baby	Robust	0.001	0.001	-21.773	(-22.796, -20.749)	81.9	(81.5, 82.2)
Baby	Robust	0.01	0.01	-24.887	(-25.996, -23.779)	80.7	(80.4, 81.1)
Baby	Robust	0.1	0.1	-45.315	(-46.451, -44.179)	68.9	(68.5, 69.3)
Baby	Robust	0.2	0.2	-66.221	(-67.343, -65.098)	59.1	(58.8, 59.5)
Tiger	Standard	0.0	0.0	20.261	(17.732, 22.790)	74.6	(74.2, 75.0)
Tiger	Standard	0.0	0.001	17.847	(15.098, 20.596)	74.6	(74.2, 75.0)
Tiger	Standard	0.0	0.01	17.549	(15.022, 20.076)	74.6	(74.2, 75.0)
Tiger	Standard	0.0	0.1	-24.003	(-28.327, -19.680)	67.4	(66.9, 67.9)
Tiger	Standard	0.0	0.2	-77.248	(-83.117, -71.379)	62.1	(61.6, 62.7)
Tiger	Robust	0.001	0.0	18.247	(15.658, 20.836)	74.8	(74.4, 75.2)
Tiger	Robust	0.01	0.0	16.593	(13.660, 19.525)	74.6	(74.2, 75.1)
Tiger	Robust	0.1	0.0	15.036	(14.287, 15.786)	81.1	(80.7, 81.5)
Tiger	Robust	0.2	0.0	-0.428	(-0.727, -0.128)	87.7	(87.3, 88.1)
Tiger	Robust	0.001	0.001	21.046	(18.606, 23.486)	74.6	(74.2, 75.0)
Tiger	Robust	0.01	0.01	18.997	(16.656, 21.338)	74.5	(74.0, 74.9)
Tiger	Robust	0.1	0.1	0.316	(-1.607, 2.238)	74.9	(74.3, 75.5)
Tiger	Robust	0.2	0.2	-13.446	(-14.238, -12.654)	74.7	(73.7, 75.7)

high. Under nominal dynamics, the robust policy has a 25% and 102% reduction in mean simulated value for ambiguity parameters 0.1 and 0.2, respectively. At lower ambiguity levels, the robust policy and nominal policy produce similar values.

The accuracy metric shows interesting trends for the Tiger Problem. The standard method produces a policy which identifies the correct state approximately 75% of the time for the nominal case. Similarly, the robust policy identifies the correct state approximately 75% of the time for the worst-case. When the nominal policy is applied to the worst-case, the accuracy decreases as the ambiguity increases, as expected. However, when the robust policy is applied to the nominal case, the accuracy increases as the ambiguity of the policy increases. This trend illustrates the conservative nature of the robust policy. While the policy updates the belief state using worst-case probabilities, the true metric improves more quickly using the nominal probabilities. As a result, the belief-state is more accurate than is rational. In other words, under nominal conditions, the robust policy spends too many resources collecting information.

The results for the Crying Baby Problem show different trends than the tiger results. In the nominal simulations, the robust solution performs significantly worse than the standard solution, as expected. However, under worst-case simulation dynamics, the robust solution is only statistically better than the nominal solution at the highest level of ambiguity and this improvement is small relative to the loss in the nominal simulations. In this case, the nominal solution is nearly robust to the amount of ambiguity considered in the experiment and the trade-off is not favorable. The robust policy has improved accuracy at higher ambiguity levels, but the improvement is not as large as in the improvement in the Tiger Problem. The difference in relative performance between the Tiger Problem and Crying Baby Problem highlights the importance of considering the trade-off between nominal-case and worst-case per-

formance.

Overall, the robust policy provides protection from undesirable dynamics, particularly in the Tiger Problem. However, the relative benefit of the worst-case protection is problem dependent in terms of both the magnitude and the likelihood of worst-case dynamics.

5.4.2 Robust Belief-Reward POMDP Experiments

In this section, we conduct an experiment to compare the belief-reward variants of the robust POMDP and POMDP. The test problem is based on the Rock Diagnosis Problem, discussed in Araya-López (2013). The full model with parameter values is defined in Appendix B.

5.4.2.1 Parameters

In addition to the parameters described in Section 5.4.1, we consider an extension to two different belief-reward functions. The first function, referred to as the simple reward function for this experiment, is a linear cone with a value of 0 at the uniform belief of $[0.25, 0.25, 0.25, 0.25]$ and value of 1 at the deterministic beliefs at each corner of the simplex. The second belief function, referred to as the complex reward function for this experiment, is defined by twice as many vectors and produces a piecewise linear convex belief function with variable gradient. The complex reward function also has a value of 0 at the uniform belief and 1 at the corners of the simplex. Figure 18 shows several belief-reward functions for a two-dimensional belief space. The simple belief-reward function in our experiment is similar to the 1-norm function, while the complex belief-reward is similar to a lower piecewise approximation of the negative entropy in the sense that it has non-constant gradient and a similar profile. We also tune the ambiguity parameters to $[0.001, 0.1, 0.2, 0.3, 0.4]$ based on the rock

diagnosis dynamics.

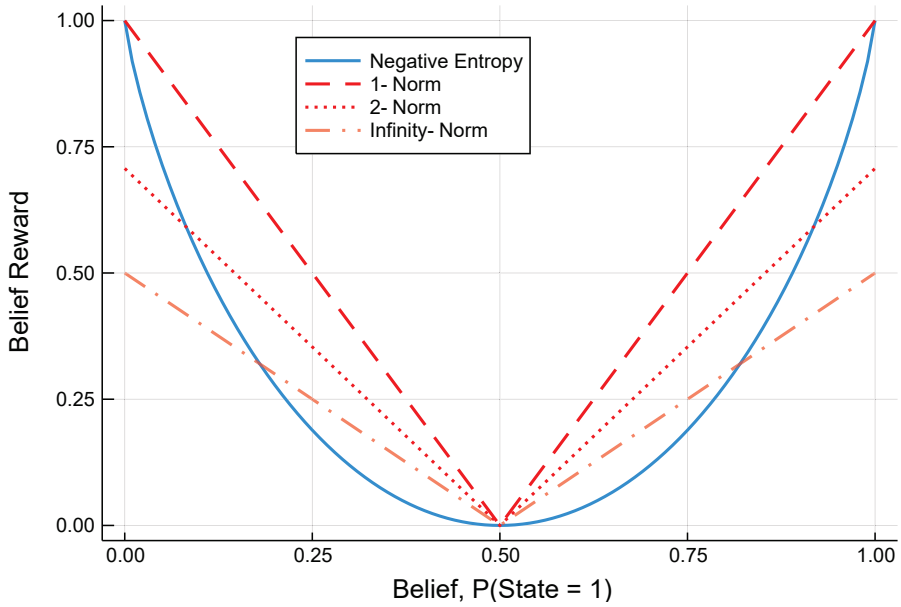

Figure 18. Belief-Reward function comparison for $|\mathcal{S}| = 2$

5.4.2.2 Results

Figures 19 and 20 show two-dimensional slices of the value functions for the nominal solution and the robust solutions to the Rock Diagnosis Problem. In this slice of the belief space, the decision maker is certain that the sensor is in position 1, which is feasible because the transition function is deterministic. The horizontal axis represents the belief that the rock is good, corresponding to a value of zero, or bad, corresponding to a value of one. In both cases, the robust policies for ambiguity size of 0.1, 0.2, 0.3, and 0.4 are similar and produce value functions that are visually indistinguishable.

For the Rock Diagnosis Problem, we fix the solution iterations at 200 with 63 uniformly spaced belief points. Then we conduct 100 replications of 100 decision epochs each to simulate the value and accuracy of each policy. The results of these tests are shown in Table 15.

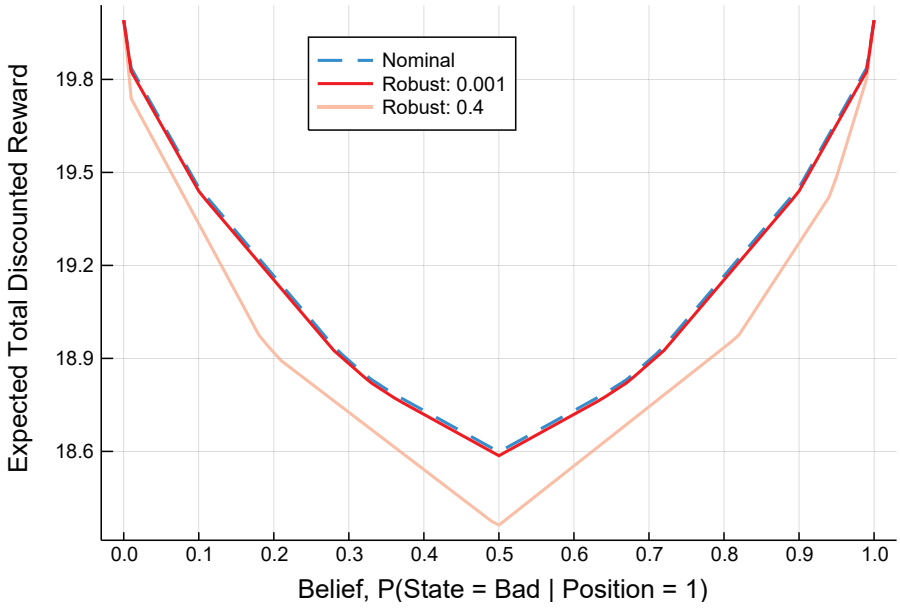

Figure 19. Robust belief-reward rock diagnosis (simple reward) POMDP value function

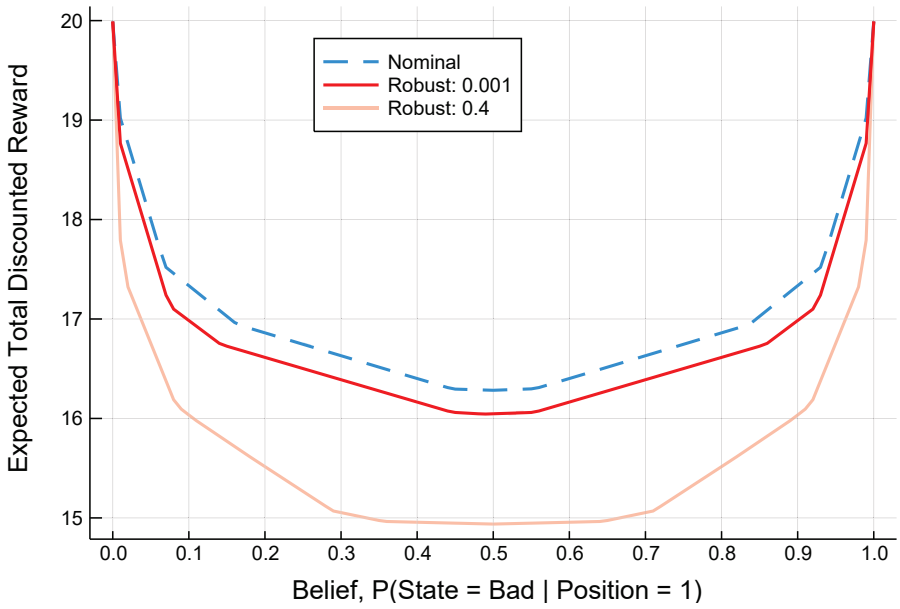

Figure 20. Robust belief-reward rock diagnosis (complex reward) POMDP value function

Table 15. Robust belief-reward POMDP and belief-reward POMDP experimental results (N = 100)

Solution	Reward	Ambiguity (Solution)	Ambiguity (Simulation)	Sim Value (Mean)	Sim Value (95% CI)	Sim % Correct (Mean)	Sim % Correct (95% CI)
Standard	Simple	0	0	18.687	(18.630, 18.743)	97.4	(95.4, 99.3)
Standard	Simple	0	0.001	18.583	(18.514, 18.652)	98.9	(98.8, 99.1)
Standard	Simple	0	0.1	18.455	(18.363, 18.547)	98.7	(98.4, 98.9)
Standard	Simple	0	0.2	18.188	(18.061, 18.315)	97.9	(97.5, 98.3)
Standard	Simple	0	0.3	17.996	(17.853, 18.140)	95.7	(94.3, 97.0)
Standard	Simple	0	0.4	17.642	(17.475, 17.809)	24.3	(17.6, 31.0)
Robust	Simple	0.001	0	18.642	(18.593, 18.691)	98.4	(98.3, 98.5)
Robust	Simple	0.001	0.001	18.607	(18.539, 18.676)	98.4	(98.2, 98.5)
Robust	Simple	0.1	0	18.411	(18.273, 18.549)	99.2	(99.1, 99.2)
Robust	Simple	0.1	0.1	18.491	(18.354, 18.628)	99.1	(99.0, 99.2)
Robust	Simple	0.2	0	18.438	(18.289, 18.587)	98	(97.8, 98.2)
Robust	Simple	0.2	0.2	18.539	(18.426, 18.652)	98.1	(98.0, 98.3)
Robust	Simple	0.3	0	18.374	(18.219, 18.529)	97.9	(97.7, 98.2)
Robust	Simple	0.3	0.3	18.463	(18.317, 18.609)	99	(98.9, 99.1)
Robust	Simple	0.4	0	18.473	(18.344, 18.602)	99	(98.9, 99.2)
Robust	Simple	0.4	0.4	18.453	(18.308, 18.597)	99	(98.9, 99.1)
Standard	Complex	0	0	16.477	(16.302, 16.652)	97.7	(97.6, 97.9)
Standard	Complex	0	0.001	16.445	(16.257, 16.633)	97.7	(97.6, 97.9)
Standard	Complex	0	0.1	16.082	(15.770, 16.394)	97.2	(94.5, 99.9)
Standard	Complex	0	0.2	14.853	(14.390, 15.317)	87.9	(82.0, 93.9)
Standard	Complex	0	0.3	14.042	(13.527, 14.556)	92.3	(89.5, 95.0)
Standard	Complex	0	0.4	13.818	(13.238, 14.398)	47.4	(38.2, 56.7)
Robust	Complex	0.001	0	16.433	(16.265, 16.601)	95.8	(93.1, 98.5)
Robust	Complex	0.001	0.001	16.358	(16.171, 16.545)	99.2	(99.2, 99.3)
Robust	Complex	0.1	0	14.654	(14.399, 14.908)	97.9	(97.6, 98.2)
Robust	Complex	0.1	0.1	14.957	(14.756, 15.159)	98.3	(98.2, 98.4)
Robust	Complex	0.2	0	16.165	(15.841, 16.490)	98	(97.8, 98.2)
Robust	Complex	0.2	0.2	14.591	(14.337, 14.846)	99.1	(98.9, 99.2)
Robust	Complex	0.3	0	14.444	(14.134, 14.754)	97.9	(97.6, 98.1)
Robust	Complex	0.3	0.3	16.183	(15.845, 16.520)	98	(96.1, 99.9)
Robust	Complex	0.4	0	15.799	(15.413, 16.184)	94.9	(91.6, 98.2)
Robust	Complex	0.4	0.4	14.273	(13.928, 14.618)	98.7	(98.5, 99.0)

In the simple belief-reward experiments, the simulated value and accuracy results for the belief-dependent setting follow the same general pattern as the results in the state-reward setting. While the nominal solution is vulnerable to degraded performance under the worst-case dynamics, the robust solution outperforms the nominal solution for high ambiguity for both metrics. Figure 21 shows this trend graphically for the simulated value metric and highlights the benefit of a robust belief-reward policy. Note that although this is a statistically significant difference, the magnitude of the difference is relatively small compared to the simulated total discounted reward for the problem.

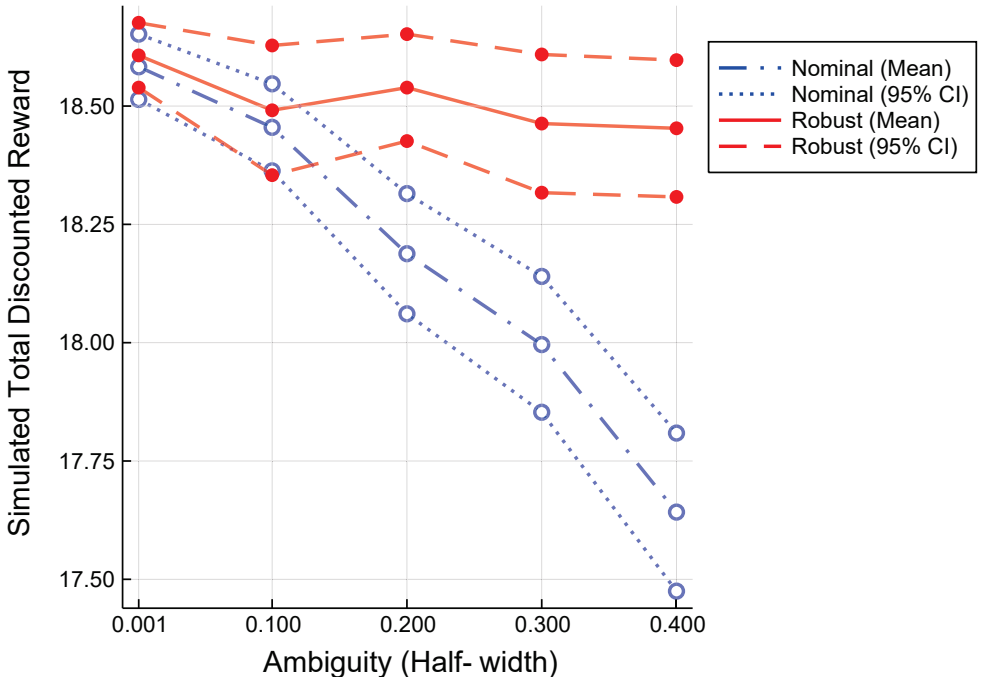

Figure 21. Simulated value for rock diagnosis (simple reward) policies with worst-case dynamics

The results for the complex reward experiments are mixed. Although the robust policy has higher simulated value than the nominal policy at the 0.3 ambiguity level with the complex reward, the difference is not as clear as in the simple reward case. For most sizes of ambiguity sets considered, the confidence intervals for the robust

policy and the nominal policy overlap. This overlap is likely due to the interaction between the shape of the belief-reward and the worst-case dynamics. However, the robust policy does outperform the nominal policy with respect to accuracy for high ambiguity levels. Similar to the difference in relative performance between the Tiger Problem and the Crying Baby Problem, the desirability of the robust belief-reward policy depends on the problem parameters. If the worst-case loss is minor compared to typical rewards, the nominal policy is preferred.

5.5 Cybersecurity Application

Cybersecurity is a critical concern in both military and civilian contexts (Schramm and Gaver, 2013; Sinha et al., 2015). In cybersecurity applications, intrusion detection systems (IDS) play an important role. In some cases, the IDS is paired with an intrusion response system (IRS) that automatically employs counter-measures for detected threats (Kiennert et al., 2018). However, the problem of detection is a difficult problem even without considering response selection. The subset of cybersecurity detection problems is closely related to sensing problems (e.g., Szechtman et al. (2008) and Romich et al. (2015)) and has also been addressed with POMDPs (Miehling et al., 2018). The IDS decision problem is a canonical example of an information-reward problem. While MDPs and POMDPs have been used to model IDS, we develop an application of the robust belief-reward POMDP to cybersecurity.

We present an application of cyber resource allocation for an IDS in this section to demonstrate the implementation of the robust framework. However, full application of the robust belief-reward POMDP method for modern IDS would require scaling to a larger problem in terms of both size and diversity of the state, action, and observation spaces to better reflect cybersecurity systems in use today. Cybersecurity systems are used widely by private industry, government, and military organizations.

In the military setting, cyber IDS are a single-domain component of more general multi-domain assessment, monitoring, and evaluation systems. Although the model described here is presented in terms of cyber assessment, the principles can be applied to other domains as well.

5.5.1 Nominal Formulation

For this application, we consider a cyber network with three nodes. For each node, the intrusion threat-level is measured on a three-level ordinal scale representing no threat (threat level 1), degraded capability (threat level 2), and critical threat (threat level 3). We also consider the IDS planning problem to be infinite-horizon with discrete-time decision epochs. IDS controllers may be limited in deployment due to cost or due to the negative impact on the usability of the network. Formally, the state space is $\mathcal{S} = \{(1,1,1),(1,1,2),\ldots,(3,3,3)\}$ where each state is represented by factored notation, $s = (s_1, s_2, s_3)$, in which the jth entry represents the threat-level of the jth node. For the second state, $s = (1,1,2)$, the first node is at threat-level 1, the second node is at threat-level 1, and the third node is at threat-level 2. Given $N_n = 3$ nodes and $N_l = 3$ threat-levels, the size of the state space is $|\mathcal{S}| = N_l^{N_n} = 27$.

For this example, we consider an IDS with $N_c = 2$ monitoring systems. The action space is the assignment of either one monitor or no monitors to each node, without doubled assignments. Then, the size of the action space is $|\mathcal{A}| = \frac{N_n!}{(N_n - N_c)!} = 6$ actions. Using a similar factored notation as the state space, the action space is $\mathcal{A} = \{(1,2),(1,3),\ldots,(3,2)\}$ where each action is represented by factored notation, $a = (a_1, a_2)$, in which the jth entry represents the node to which the jth monitor is assigned. For example, $a = (1,2)$ assigns the first monitor to node 1 and the second monitor to node 2.

The observation space captures the threat-level detected by each monitor. Each

monitor can detect any of the three threat-levels, so the size of the observation space is $|\mathcal{Z}| = N_l^{N_c} = 9$. Using the same factored approach, the observation space is $\mathcal{Z} = \{(1,1),(1,2),\ldots,(3,3)\}$ where each observation is represented by factored notation, $z = (z_1, z_2)$, in which the jth entry represents the threat-level which the jth monitor detects. For example, if $z = (1,3)$, the first monitor detects threat-level 1 and the second monitor detects threat-level 3.

The transition function describes how the state changes based on the actions, in other words, the dynamics of the problem. In this case, the monitoring actions are assumed to have no impact on the state, so the transition function is relatively simple. The transition function is also assumed to be the same for each node. The state transition function for this scenario represents the *a priori* likelihood of the state transitioning, given suspected threats.

There are six parameters that control transitions in this model, shown in Table 16. These probabilities represent the likelihood of independent events of a given node changing threat-levels in one decision epoch, where s_i and s_i' indicate the threat-level of node i before and after the decision epoch and l and k are threat-levels. Note that these constants are defined with respect to a single node, then used to determine the probability for the full state that includes all three nodes. The basic structure is that a given node will stay at the same threat-level, decrease one threat-level, or increase one threat-level without skipping over levels. When a node is at the highest or lowest threat-level, the distribution has to be modified to prevent transitions to threat-levels outside the state space. We assume that all nodes have the same probability of changing, but that assumption can be generalized in a straightforward manner by changing the value of the parameters. Furthermore, complexity increases for larger systems where one node's transitions affect the transition probabilities of neighboring nodes.

Table 16. Nominal transition parameters

Parameter	Value	Definition	Explanation
p_d	0.15	$p(s'_i = l-1 \mid s_i = l)$, $i \in \{1,2,3\}, l > 1$	probability of decrease
p_u	0.3	$p(s'_i = l+1 \mid s_i = l)$, $i \in \{1,2,3\}, l < 3$	probability of increase
p_s	0.55	$p(s'_i = 2 \mid s_i = 2)$, $i \in \{1,2,3\}$	probability of no change
$p_{s\cup d}$	0.7	$p(s'_i = 1 \mid s_i = 1)$, $i \in \{1,2,3\}$	probability of no change (lower limit)
$p_{s\cup u}$	0.9	$p(s'_i = 3 \mid s_i = 3)$, $i \in \{1,2,3\}$	probability of no change (upper limit)
p_h	0	$p(s'_i = k \mid s_i = l)$, $i \in \{1,2,3\}$, $k \notin \{l-1, l, l+1\}$	probability of skipping levels

These constants and the dynamics discussed previously define the transition array, $\mathbf{T}$. The full transition array is a $27 \times 6 \times 27$ array, but each two-dimensional slice for a given action is the same because the transition does not depend on the action selected. For notational clarity, we give the state s an integer-valued index, k_s, numbered in increasing ternary order, as in the definition for S. For convenience, we refer to state s by its identifying variable and also by its index, using context to distinguish. For example, the first state, $(1,1,1)$, has index $k_s = k_{(1,1,1)} = 1$, so we will refer to it both as state 1 and as $(1,1,1)$. We treat the notation for actions and observations in an analogous manner. Formally, the transition function can be defined as $T(s,a,s') = \mathbf{T}_{k_s,k_a,k_{s'}}$, where $\mathbf{T}_{k_s,k_a,k_{s'}}$ is the $(k_s,k_a,k_{s'})$ entry of the $\mathbf{T}$ array. All transition matrices indexed by action are available in the source code.

Since the observation function does depend on the actions, unlike the transition function, we have a three-dimensional array, with dimensions $6 \times 27 \times 9$. Each two-dimensional slice in this array represents the observation function given a certain action. That is, we have an array $\mathbf{O}$, where entry (i,j,k) indicates the probability of observing state k given the current state is j and the action chosen is i. Then

$O(a, s', z) = \mathbf{O}_{k_a, k_{s'}, k_z}$.

For the detection process of collecting information and determining the threat-level of a given node, we set a performance parameter for each monitor, where $p^1_{\text{correct}} = 0.8$ is the probability of correct detection of the first monitor and $p^2_{\text{correct}} = 0.9$ is the probability of correct detection of the second monitor. The probability of incorrect detection is uniformly distributed over the adjacent incorrect threat-levels, but zero for threat-levels with more than one level of error. For instance, if the actual state of node 1 is threat-level 2 and we assign monitor 1 to node 1, then we will observe node 1 at threat-level 1 with probability 0.1, threat-level 2 with probability 0.8, and threat-level 3 with probability 0.1. We assume independent observations for different nodes. For a given node, if no monitor is assigned, no change in its threat-level can be observed. These parameters are summarized in Table 17, where z_i is the observation of monitor i and a_i is the node to which monitor i is assigned. Similar to the transition table, we define the observation parameters with respect to a single node rather than the entire state. The difficulty in estimating these probabilities in practice is one of the primary reasons for incorporating ambiguity in the robust formulation. All observation matrices indexed by action are also available in the source code.

The belief-reward is $\rho(\mathbf{b}) = \max_{\alpha \in \Gamma} \mathbf{b}\alpha$, defined using alpha-vectors over the state space. We use a simple belief-reward function defined as a set of vectors, Γ, that includes $|S| = 27$ vectors. Each vector has a value of 1 at a single position and $-1/26$ otherwise, giving a value of 1 at each corner of the belief simplex and a value of 0 at the center of the simplex. The value increases linearly toward each extreme point of the belief simplex, creating a PWLC cone. Formally, the alpha vector set is

$$\Gamma = \{[1, -1/26, -1/26, \ldots, -1/26], \tag{32}$$

$$[-1/26, 1, -1/26, \ldots, -1/26], \tag{33}$$

$$\vdots \tag{34}$$

$$[-1/26, -1/26, \ldots, -1/26, 1]\}. \tag{35}$$

Table 17. Nominal observation parameters

Parameter	Value	Definition	Explanation
p^1_{correct}	0.8	$p(z_1 = l \mid s_{a_1} = l)$, $l \in \{1, 2, 3\}$	probability monitor 1 correct
p^2_{correct}	0.9	$p(z_2 = l \mid s_{a_2} = l)$, $l \in \{1, 2, 3\}$	probability monitor 2 correct
q^1_{center}	0.1	$p(z_1 = l \mid s_{a_1} = 2)$, $l \in \{1, 3\}$	probability monitor 1 incorrect
q^2_{center}	0.05	$p(z_2 = l \mid s_{a_2} = 2)$, $l \in \{1, 3\}$	probability monitor 2 incorrect
q^1_{boundary}	0.2	$p(z_1 = 2 \mid s_{a_1} \in \{1, 3\})$	probability monitor 1 incorrect (boundary)
q^2_{boundary}	0.1	$p(z_2 = 2 \mid s_{a_2} \in \{1, 3\})$	probability monitor 2 incorrect (boundary)
p_0	0	$p(z_i = k \mid s_{a_i} = l)$, $i \in \{1, 2\}$, $k \notin \{l-1, l, l+1\}$	probability of two-level error

We set the initial state to $(1, 1, 1)$ so that the network starts with no threats present, and we set $\gamma = 0.9$ to a moderate discount rate to reflect the dynamic nature of the security environment.

The nominal belief-reward POMDP model for cyber intrusion detection planning is the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{Z}, T, O, \rho, \gamma)$:

$$\mathcal{S} = \{(1,1,1), (1,1,2), \ldots, (3,3,3)\} \tag{36}$$

$$\mathcal{A} = \{(1,2), (1,3), \ldots, (3,2)\} \tag{37}$$

$$\mathcal{Z} = \{(1,1), (1,2), \ldots, (3,3)\} \tag{38}$$

$$T(s, a, s') = \mathbf{T}_{k_s, k_a, k_{s'}} \tag{39}$$

$$O(a, s', z) = \mathbf{O}_{k_a, k_{s'}, k_z} \tag{40}$$

$$\rho(\mathbf{b}) = \max_{\alpha \in \Gamma} \mathbf{b}\alpha \tag{41}$$

$$\gamma = 0.9 \tag{42}$$

5.5.2 Robust Formulation

Using the nominal dynamics, we define the robust dynamics with probability intervals. For the ambiguity in the transition and observation functions, we can introduce ambiguity into the transition array directly or into the parameters used to calculate the dynamics. Introducing ambiguity into the parameters seems to be a more natural way to represent real world ambiguity, so that approach is used here. To facilitate interpretation, we vary the size of the ambiguity sets but keep the transition function near certain for all distributions. For the observation function, we consider the performance of one monitor to be near certain while the other has high ambiguity. Specifically, we set a one-sided ambiguity parameter for the transitions, $\delta_t = 1\text{e}{-6}$, and for each monitor, $\delta_1 = 1\text{e}{-6}$ and $\delta_2 = 0.4$. These ambiguity parameters are added and subtracted to each nominal parameter used to generate the transition and observation arrays. We bound any interval by ϵ and $1 - \epsilon$ if the raw ambiguity interval extends beyond probability bounds. This approach generates an array of upper probabilities and an array of lower probabilities, neither of which have rows that sum to one. The probability intervals defined by these two arrays are used to define a closed, convex polytope that contains valid probability distributions.

For the transition function, we apply the δ_t ambiguity to the three main parameters, p_d, p_s, and p_u in Table 16, and derive the ambiguity for the union parameters. We keep the probability of skipping levels precise at 0. The interval widths of the resulting array are near zero, modeling a high level of certainty about the transition dynamics. Using the same logic as the nominal model dynamics, we calculate an upper transition array using the upper bounds and a lower transition array using the lower bounds. These arrays represent the upper and lower probabilities of transition, so we have two arrays with real-valued elements rather than a single array with interval-valued elements. We denote the upper and lower transition arrays by $\mathbf{T}^U$

and $\mathbf{T}^L$. Given these arrays, and the element indexing discussed earlier, the robust transitions are defined by $T^R(s,a,s') = [\mathbf{T}^L_{k_s,k_a,k_{s'}}, \mathbf{T}^U_{k_s,k_a,k_{s'}}]$. The full upper and lower transition arrays are available in the source code.

The observation array is developed in an analogous way. We use δ_1 and δ_2 to create the interval-valued parameters shown in Table 18. The logic for the observation function remains the same as in the nominal case and when applied to both end points of the parameter intervals, this produces upper and lower observation arrays $\mathbf{O}^U$ and $\mathbf{O}^L$. Given these arrays, the robust observations are defined by $O^R(a,s',z) = [\mathbf{O}^L_{k_a,k_{s'},k_z}, \mathbf{O}^U_{k_a,k_{s'},k_z}]$. This robust observation interval-array models low ambiguity about the first monitor and high ambiguity about the second monitor.

Table 18. Robust observation parameters

Parameter	Value
p^1_{correct}	$[0.8 - 1\text{e}{-6}, 0.8 + 1\text{e}{-6}]$
p^2_{correct}	$[0.5, 1 - 1\text{e}{-6}]$
q^1_{center}	$[0.1 - 5\text{e}{-7}, 0.1 + 5\text{e}{-7}]$
q^2_{center}	$[5\text{e}{-7}, 0.25]$
q^1_{boundary}	$[0.2 - 1\text{e}{-6}, 0.2 + 1\text{e}{-6}]$
q^2_{boundary}	$[1\text{e}{-6}, 0.5]$
p_0	0

The belief-reward and discount factor remain unchanged from the nominal setting. To summarize, the robust belief-reward model is the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{Z}, \mathcal{P}, \rho, \gamma)$ where $\mathcal{P}$ is defined by T^R and O^R, and where

$$\mathcal{S} = \{(1,1,1), (1,1,2), \ldots, (3,3,3)\} \tag{43}$$

$$\mathcal{A} = \{(1,2), (1,3), \ldots, (3,2)\} \tag{44}$$

$$\mathcal{Z} = \{(1,1), (1,2), \ldots, (3,3)\} \tag{45}$$

$$T^R(s,a,s') = [\mathbf{T}^L_{k_s,k_a,k_{s'}}, \mathbf{T}^U_{k_s,k_a,k_{s'}}] \tag{46}$$

$$O^R(a,s',z) = [\mathbf{O}^L_{k_a,k_{s'},k_z}, \mathbf{O}^U_{k_a,k_{s'},k_z}] \tag{47}$$

$$\rho(\mathbf{b}) = \max_{\alpha \in \Gamma} \mathbf{b}\alpha \tag{48}$$

$$\gamma = 0.9 \tag{49}$$

5.5.3 Results

We use standard belief-reward value iteration to solve the cybersecurity problem for the nominal formulation and robust belief-reward value iteration to solve the robust formulation. For comparison purposes, we also consider an off-nominal formulation with no ambiguity but different observation parameters than the nominal formulation. Specifically, the off-nominal formulation has $p^1_{\text{correct}} = 0.8$ and $p^2_{\text{correct}} = 0.5$.

After computing the solution for each formulation, we test each policy using an empirical simulation in which the transitions and observations are generated from the worst-case distribution. The beliefs and rewards are generated using the dynamics associated with each policy. We also test each policy using an empirical simulation in which the transitions and observations are generated from the nominal distribution.

The performance of each policy is shown in Tables 19 and 20. These results were computed using 12 selected belief points and 40 uniformly random points in the belief space. The selected belief points are chosen from points near the starting belief and are used to check results during the simulation. Following the approach in Pineau et al. (2003), the solution policy was calculated using 45 iterations of value-iteration as $|\rho_{\max} - \rho_{\min}|\gamma^{N_{\text{iter}}} = 0.9^{45} < 0.01$. For the simulation, 25 replications of 100 steps each were performed. In production point-based value iteration algorithms, the belief set is dynamically adjusted, typically by balancing exploration and exploitation with reachable beliefs (e.g., Smith and Simmons (2005); Kurniawati et al. (2008)). However, in this case, we preferred to test the robust belief-reward point-based value iteration against a static belief set to avoid an artificial disadvantage for the nominal case. If beliefs not reachable under nominal conditions are not used to develop the

nominal policy, it would perform worse than it otherwise would in the off-nominal and robust settings.

Table 19. Empirical cybersecurity belief accuracy ($N = 25$)

Policy	Nominal		Worst-Case	
	Accuracy (Mean)	Accuracy (95% CI)	Accuracy (Mean)	Accuracy (95% CI)
Standard (Nominal)	51.1	(47.8, 54.4)	14.1	(12.6, 15.7)
Standard (Off-Nominal)	38.7	(35.1, 42.3)	16.0	(13.7, 18.3)
Robust	41.0	(38.7, 43.2)	22.8	(20.5, 25.1)

Table 20. Empirical cybersecurity belief-reward ($N = 25$)

Policy	Nominal		Worst-Case	
	Reward (Mean)	Reward (95% CI)	Reward (Mean)	Reward (95% CI)
Standard (Nominal)	5.06	(4.93, 5.19)	4.68	(4.51, 4.84)
Standard (Off-Nominal)	4.10	(3.98, 4.21)	3.76	(3.63, 3.89)
Robust	3.87	(3.73, 4.01)	3.93	(3.79, 4.07)

As expected, the robust policy has the lowest simulated total discounted reward when applied to the nominal simulation in Table 20. However, when applied to the worst-case simulation, the simulated total discounted reward of the robust policy remains stable while the simulated total discounted rewards of the nominal and off-nominal policies decrease but remain higher than that of the robust policy. Table 19 suggests that this finding is a result of over-confidence for the precise policies under model misspecification. Unlike in the standard reward setting, in the belief-reward setting, the rewards in the simulation are a function of both the simulation dynamics and solution dynamics, because the rewards depend on the policy's belief update. Although the nominal policy has beliefs with relatively high certainty, even in the worst-case simulation, the accuracy metric shows that the actual performance of the nominal policy is worse than the performance of the robust policy, significant at $p = 0.05$. In other words, the nominal policy reports inaccurate but highly certain beliefs,

due to the model misspecification. Over time the conflict between the inaccurate belief and the true dynamics negatively impacts the belief-reward but the impact is sensitive to the discount parameter and the model structure. As a result, the accuracy metric is a better measure of performance for each policy for this application.

For this cybersecurity assessment problem, the robust policy has a 61% improvement in the probability of correctly identifying the state of the defended network under the worst-case problem dynamics. This improvement comes at a cost of a 20% decrease in the probability of correctly identifying the state of the defended network under nominal conditions. In terms of raw percentage points, the relative improvement in the worst-case is 86% of the magnitude of the loss in the nominal case. Despite high probabilities of correct detection for each monitor, the likelihood of correctly identifying the state of the network, even under nominal conditions, is low for all three policies. The difficulty in correctly identifying the state of the network is related to the high dimensionality introduced the by the combinatorial nature of the nodes and threat-levels. Overall, the robust policy provides improved performance under worst-case problem dynamics at a reasonable cost in the nominal case.

5.5.4 Sensitivity Analysis

In this section, we conduct an experiment to show that the improvement due to the robust approach is not restricted to the particular parameter values used in Section 5.5.2. The three key parameters that we vary in this experiment are the relative state transition probabilities, the relative accuracy of the two intrusion detection systems (IDSs), and the relative ambiguity between the two IDSs. We parameterize the relative state transition probabilities by changing the value of the probability of no change, p_s, and keeping the ratio of the probability of increase, p_u, to the probability of decrease, p_d, at $2:1$. For the remaining two parameters, we fix p_2 and δ_1 and vary the

relative accuracy and ambiguity. The relative accuracy is $p^2_{\text{correct}} - p^1_{\text{correct}}$ and the the relative ambiguity is $\delta_2 - \delta_1$. We fix the alternate ambiguity and accuracy parameters to avoid a situation in which one sensor strictly dominates on both accuracy and ambiguity. Using these three parameters, we build a central composite design with six center points and three total replications (Myers et al., 2016). The response for each run includes the nominal belief-reward, the worst-case belief-reward, the nominal accuracy, and the robust accuracy for both the nominal policy and the robust policy. This experimental design results in 60 total runs with design settings summarized in Table 21. We select a broad range of parameter values to explore the full range of reasonable parameters for this application area.

Table 21. Cybersecurity experimental design

	Value				
Factor	Lower Axial	Lower	Center	Upper	Upper Axial
Transition Parameter	0.37	0.4	0.5	0.6	0.63
Accuracy Parameter	0.04	0.1	0.3	0.5	0.56
Ambiguity Parameter	0.04	0.1	0.3	0.5	0.56

The results of this experiment are shown in Figures 22 and 23. The overall response surface model has an adjusted R^2 value of 0.74, with the noise due to the inherent variability in the simulation. The model includes four significant effects. The ambiguity main effect and second order effect are both significant at the 0.05 level with p-values less than 0.001 and practically meaningful parameter estimates (5.7 and 3.1, respectively). The second-order effects for the transition and IDS accuracy parameters are statistically significant at the 0.05 level with p-values of 0.005 and 0.035, respectively. However, the parameter estimates for these effects are about one third the magnitude of the ambiguity effect (1.9 and -1.4, respectively). No other effects are significant.

As in the base-case, we find that the robust policy outperforms the nominal policy with respect to IDS accuracy in the worst-case across a broad range of parameter

settings. Additionally, the nominal policy shows incorrectly high performance with respect to belief-reward across the range of parameter settings. Figure 22 shows the improved accuracy of the robust policy using percent accuracy change, defined as the increase in accuracy from the nominal policy to the robust policy in terms of raw percentage change. The improved accuracy of the robust policy is statistically significant at moderate ambiguity levels and increases as ambiguity increases.

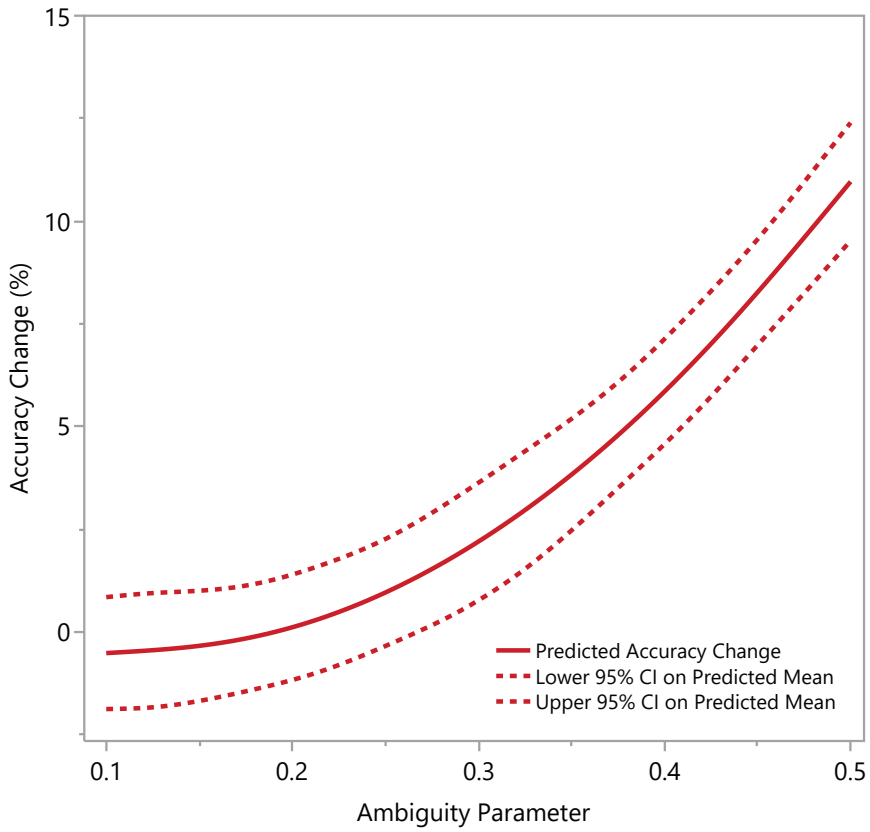

Figure 22. Cybersecurity percent accuracy change by ambiguity level (robust policy - nominal policy)

Furthermore, the improvement in accuracy for the robust policy holds regardless of the value of the transition parameter or the IDS accuracy parameter, as seen in Figure 23. In the upper-right cell of Figure 23, the relationship between accuracy change and ambiguity level is similar when the transition parameter is set to 0.4 and

0.6. Similarly, in the center-right cell of Figure 23, the relationship between accuracy change and ambiguity level is similar when the accuracy parameter is set to 0.1 and 0.5. The relationship between accuracy change and ambiguity level is also similar in both of these cells. These qualitative results confirm the small interaction effects found in the statistical response surface model.

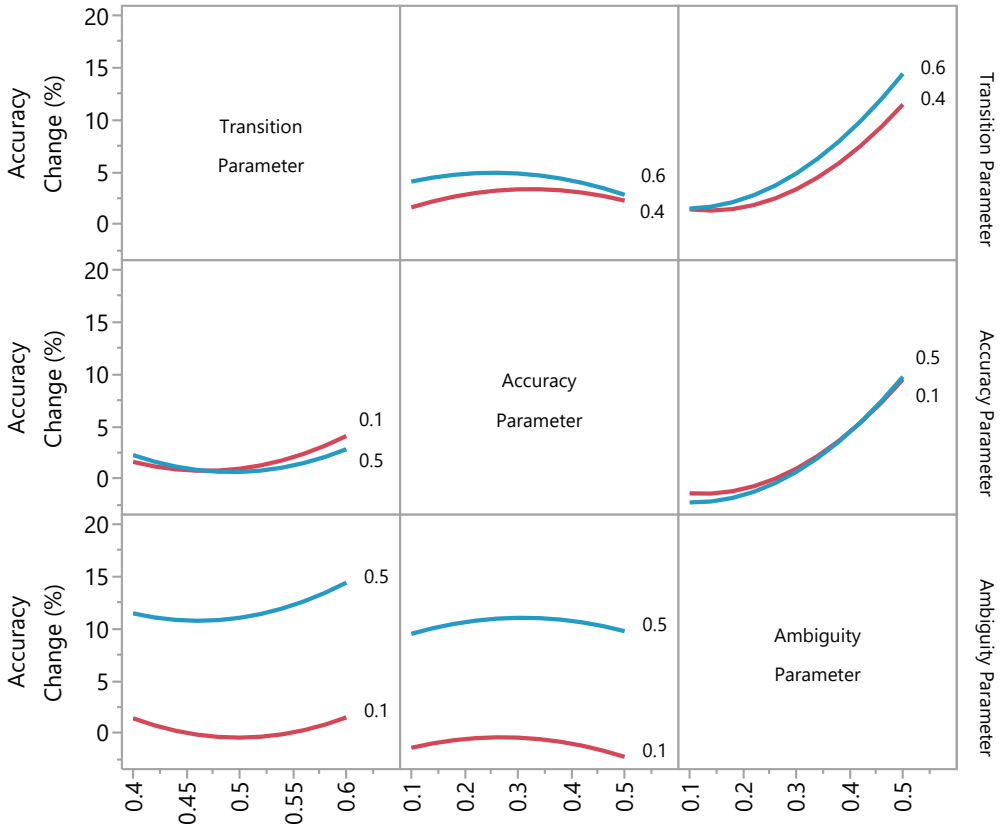

Figure 23. Sensitivity of cybersecurity accuracy change to transition, accuracy, and ambiguity parameters

5.6 Conclusion

In this chapter, we develop a comprehensive approach to address the challenge of POMDP model misspecification. This approach includes a formulation for robust belief-reward POMDPs as a generalization of standard MDPs that allows for state

uncertainty, model uncertainty, and belief-dependent rewards. To solve these models, we show how to extend approximate solution techniques using the new robust belief-reward point-based value iteration algorithm. This algorithm has desirable convexity and convergence properties that make it an efficient approximate solution technique while still addressing the general case of ambiguous dynamics.

We also provide experimental results for robust POMDPs and robust belief-reward POMDPs. The experimental results for robust belief-reward POMDPs show that the robust belief-reward value iteration algorithm outperforms standard belief-reward value iteration under worst-case model dynamics. This worst-case protection is valuable for applications with ambiguous transition and observation dynamics, which are common in practice.

Finally, we present a detailed application of the robust belief-reward formulation and algorithm to a cybersecurity resource allocation problem. In this setting, we show that model misspecification can be particularly problematic, resulting in an overconfident nominal solution. The robust policy avoids overconfidence and achieves higher accuracy intrusion detection than the nominal policy. This new approach enables future cybersecurity analysis applications that address realistic uncertainty in the domain.

This work also provides a foundation for future work addressing misspecification in POMDPs. A natural progression from these results is to introduce further modifications to the approximation technique to address larger problems. Current optimal POMDP solution methods are still exponential in the observation space (Smith and Simmons, 2005; Walraven and Spaan, 2017). Recent offline solution techniques use a point-based approach with alternating lower and upper bound updates to find tight bounds on solutions for problems on the order of 10^5 states (Smith and Simmons, 2004; Kurniawati et al., 2008). Recent online techniques find high quality approxi-

mate solutions for problems of the same size and can also solve larger problems on the order of 10^{56} states and 10^3 observations (Ye et al., 2017). Applying the robustness results from this chapter to large-scale POMDP approximate solution techniques provides the ability to expand the practical applications by reducing computational time for large-scale problems while still addressing model misspecification.

This work focuses on the value iteration family of solution techniques, but the policy iteration family might also be useful. Policy iteration can be more efficient than value iteration for standard POMDPs (Hansen, 1997). Recent results have extended modified policy iteration to the robust MDP setting (Kaufman and Schaefer, 2013). This approach is a useful foundation for further extensions to develop a robust modified policy iteration algorithm for POMDPs in addition to the robust value iteration approach proposed here.

As POMDP applications continue to address larger and more difficult problems, model misspecification is an increasingly important risk factor for implementation, especially in cyber applications. The robust belief-reward POMDP formulation and algorithm presented here provide a principled approach for developing policies for robust sequential decision making in uncertain and ambiguous environments.

VI. Robust Multi-Agent Sequential Optimization Under Uncertainty

6.1 Introduction

In this chapter, we develop a game-theoretic strategy to optimally defend a set of high value assets from an integrated cyber-physical threat. This type of resource allocation in an adversarial security environment is a *security game*. Security games address resource allocation for adversarial problems in a variety of domains, including drug interdiction, weapons trafficking, illicit finance, wildlife protection, forestry protection, fishery protection, urban crime, missile defense, and cybersecurity (Kar et al., 2016). This research area uses game theory, planning under uncertainty, and optimization techniques to find optimal and approximate security policies, which are often implemented in practice (Kar et al., 2016). For instance, LAX airport has used the ARMOR program for managing security checkpoints and the US Federal Air Marshals have used the IRIS program for scheduling flights (Korzhyk et al., 2011).

As a result of the security game literature's focus on practical effectiveness, there are two critical properties for solution techniques in this area. First, solution techniques need to scale to address the full scope of practical protection and interdiction problems, for instance over large geographical areas, such as fishery protection in the Gulf of Mexico (Haskell et al., 2014) and checkpoint location in Mumbai (Jain et al., 2013). Second, solution techniques need to address realistic adversaries. This requirement has motivated behavioral models, robust models, and robust behavioral models (Nguyen et al., 2014, 2016). These models address the bounded rationality of real world adversaries by departing from classic Nash equilibrium solution concepts. Recent attention in the operations research literature has expanded the available robust game theoretic formulations, including robust games (Aghassi and Bertsimas,

2006), robust stochastic games (Kardeş et al., 2011; Kardeş, 2014), and distributionally robust games (Qu et al., 2017; Singh et al., 2017; Liu et al., 2018). However, the operations research literature has not addressed behavioral adversaries as thoroughly.

The security game literature also draws on economics and computer science to model adversaries, or *opponents*, with bounded rationality. The robust approach takes a pessimistic view that attempts to exploit an opponent's bounded rationality while exposing the decision maker to controlled risk of exploitation. There is a fundamental tension between Nash equilibrium strategies that are guaranteed to produce good results and best-response strategies that *exploit* sub-optimal play by the opponent. Strategies that exploit weak play generally produce higher rewards against a weak opponent than Nash equilibrium strategies produce against a weak opponent. However, these exploitation strategies are also generally vulnerable to exploitation by a sophisticated adversary that temporarily imitates a weak opponent. This trade-off between the *safe* Nash equilibrium strategy and the risky exploitation strategy mirrors the risk-reward trade-off between robust policies and expectation-maximization policies in traditional optimization.

Cybersecurity has been addressed through both the security literature (Sinha et al., 2015) and the cyber domain literature (Kiennert et al., 2018). In addition to the standard Stackelberg security game formulation, cybersecurity problems are often modeled as stochastic games (Shen et al., 2007; Chen and Leneutre, 2009), including partially observable stochastic games (Zonouz et al., 2014). In addition to heuristic tree searches, such games can be solved with approximate value iteration approaches (Hansen et al., 2004; Kumar and Zilberstein, 2009; Horak et al., 2017; Ye et al., 2017) or non-linear programming approaches (Wray et al., 2018).

From the physical perspective, missile defense security games have been addressed as weapon-target assignment problems (Kline et al., 2018), location and dispatching

problems (Han et al., 2016; Davis et al., 2017; Boardman et al., 2017) and sensor management problems (Lessin et al., 2018). The classic weapon-target assignment approach focuses specifically on missile-to-missile engagement (Ahner and Parson, 2014; Kline et al., 2017). Missile defense is also closely related to the broader security game and network security literature (Nystrom et al., 2018; Alpern et al., 2011).

In the typical approach of the security game literature, particularly in Stackelberg security games, the defender commits to a strategy which the attacker then observes before attacking. However, in many realistic security scenarios, the attacker does not have perfect information. Furthermore, multi-domain operations and multi-domain command and control are emerging as critical focus areas in security applications. From a multi-domain perspective, it is important to integrate both the physical missile defense and cyber missile defense together while accounting for limited defender and attacker knowledge. In military applications, there is often a significant time delay between the defender's implementation of physical security measures and the attacker's physical attack. In the intervening period, the attacker is likely to deploy cyber attacks against the defender's physical assets. This research advances the security game literature by addressing an integrated and multi-period air and cyber threat with imperfect information, and by allowing for continuous, probabilistic strategies. This work also presents the first comparison of Monte Carlo, discounted, and robust counterfactual regret algorithms for security games.

The remainder of the chapter is organized as follows. In Section 6.2, we introduce the problem formulation for the multi-domain security game. In Section 6.3, we describe the application of solution methodologies for optimal and approximate algorithms. In Section 6.4, we present exact and approximate computational results for a realistic problem. We also show the impact of different levels of cyber situational awareness, investigate the impact of stepsize on the convergence rate of the approxi-

mate algorithm, and explore a broad range of problem-specific parameter settings. In Section 6.5, we present an extension to robust exploitation of opponents with bounded rationality. Finally, in Section 6.6, we summarize the results and suggest areas for future research.

6.2 Problem Formulation

6.2.1 Scenario

In this section, we introduce a *multi-domain security game* for national defense. This game is an extensive form, zero-sum game with two players, an attacker and a defender, which have perfect recall of their actions and observations. The defender's overall objective is to protect a set of population centers from a physical attack by the adversary. In addition to the traditional physical domain, we address the cyber domain by considering a nested cyber attacker-defender game within the traditional physical security game. At a high level, we have four sequential stages: the defender's physical allocation, the attacker's cyber targeting, the defender's cyber allocation, and the attacker's physical targeting. Between each player's decision, a special player representing *nature* also acts stochastically. This game can be solved as a single extensive form game with imperfect information. However, due to the difference in time scales across domains, we consider the initial physical defensive structure to be fixed while the cyber decisions and the final attack decision are more flexible.

Specifically, we consider a nested attacker-defender, attacker-defender problem with the parameters shown in Table 22. This problem has $|\mathcal{M}|$ population centers to defend by locating air defense systems in a subset of the population centers, $\mathcal{D} \subseteq \mathcal{M}$. The population centers are vulnerable to physical attack, as in previous literature, but we also consider the air defense systems to be vulnerable to cyber attack.

To model the physical components, we consider population centers, physical de-

Table 22. Integrated cyber and air defense parameters

$\mathcal{M}$	Set of population centers
$\mathcal{D}$	Set of air defense locations
r	Coverage radius for each air defense asset
n_d	Number of air defense assets
n_a	Number of air attack targets
$\tilde{m}_d$	Number of cyber defense-capable air defense assets
$\tilde{m}_a$	Number of cyber attack-vulnerable air defense assets
$\tilde{n}_d$	Number of cyber defense teams
$\tilde{n}_a$	Number of cyber attack targets
p	Probability of effective physical defense
$\tilde{p}_s$	Probability of cyber sensor detection
$\tilde{p}_d$	Probability of effective cyber defense
$\tilde{p}_a$	Probability of effective cyber attack

fensive systems, and physical offensive systems. There are n_d physical integrated air defense systems (IADS) and n_a population centers targeted by the attacker. Each air defense system is a physical defensive system with an effective radius, r. Population centers, or cities, covered by an operational air defense system's effective radius are defended with probability p, if attacked. As modern air defense systems incorporate a variety of physical defenses, we model at the system level rather than at the individual missile level.

To model the cyber component, we consider automated cyber defenses, cyber defense teams, and cyber attacks. We assume there are automated cyber defense assets at each IADS node, $\tilde{n}_d$ human cyber defense assets, and $\tilde{n}_a$ human cyber attack assets. The automated defensive component includes both intrusion detection systems (IDS) and intrusion response systems (IRS) (Miehling et al., 2018). In the ballistic missile defense context, these automated cyber defenses have been modeled as a system that includes automated cyber sensors, analyzers, decision mediators, and actuators, in addition to human decision makers (Gagnon et al., 2010). While the automated IDS and IRS can respond to some cyber attacks, other more complex attacks require human input. United States Cyber Command's Cyber Mission Forces

achieved full operational capability in 2018, with their mission including defending the nation "by identifying adversary activity, blocking attacks and maneuvering to defeat them" (United States Cyber Command, 2018). Furthermore, according to the United States Air Force Strategic Master Plan (2015b), effective security strategy requires that defense forces "integrate and employ capabilities operating in or through the cyberspace and space domains in addition to air capabilities." Although recently activated, these highly trained Cyber Mission Forces are a limited resource and are not always available, unlike automated cyber protection systems. Cyber attack assets are also a limited resource and gaining access to each adversary system requires unique effort.

The cyber component also includes several key probabilistic outcomes. Over time, cyber vulnerabilities are discovered and patched. The lag between the attacker's discovery and the defender's awareness and deployment of a patch introduces a vulnerability to the cyber system. During this lag, the attacker has access to cyber exploits, but the availability of these capabilities are difficult for both the attacker and defender to predict. We model this cyber attack capability as a uniformly random assignment of access to a subset of the defender's IADS nodes, representing a new exploit unknown to the defender. This subset has size $\tilde{m}_a$. Similarly, it is difficult for the defender to predict how effectively it will be able to respond to future vulnerabilities, so we model the defender's cyber defense capabilities as a uniformly random subset of the defender's IADS nodes which are eligible for Cyber Mission Force defense. This subset has size $\tilde{m}_d$. The relative effectiveness of the cyber exploit, $\tilde{p}_a$, and the cyber defense, $\tilde{p}_d$, determine how likely it is a cyber defense team can neutralize a cyber attack. Additionally, each cyber attack may or may not be detected by the automated cyber IDS, determined by the probability of detection, $\tilde{p}_s$. We separate the probabilistic nodes where Nature acts into several different stages to facilitate

modeling the asymmetric information between the defender and the attacker.

Given these physical and cyber parameters, the defender's objective is to protect the population centers with payoffs defined by the expected loss of life. The attacker's objective is to maximize loss of life by allocating cyber and physical attack resources, consistent with the zero-sum formulation. To summarize, we have the game structure shown in Table 23.

Table 23. Integrated cyber and air defense game structure

Stage	Domain	Player	Size	Action
0	Physical	Defender	1	Fixed allocation of air defense assets to cities
1	Cyber	Nature	$\binom{\lvert\mathcal{M}\rvert}{\tilde{m}_d}$	Select subset of cyber defense capable nodes
2	Cyber	Nature	$\binom{\lvert\mathcal{M}\rvert}{\tilde{m}_a}$	Select subset of cyber attack capable nodes
3	Cyber	Attacker	$\binom{n_d}{\tilde{n}_a}$	Attack subset of air defense systems
4	Cyber	Nature	$2^{\tilde{n}_a}$	Determine effectiveness of cyber sensors
5	Cyber	Defender	$\binom{n_d}{\tilde{n}_d}$	Assign cyber teams to subset of air defenses
6	Physical	Attacker	$\binom{\lvert\mathcal{M}\rvert}{n_a}$	Attack subset of cities

6.2.2 Model

Given the imperfect-information extensive form structure, we follow the notation introduced by Shoham and Leyton-Brown (2008) to develop the sequence-form representation of the game. We define $G = (\mathcal{N}, \Sigma, g, \mathcal{K})$, where

$\mathcal{N} = \{1, 2\}$ is the set of agents,

$\Sigma = (\Sigma_1, \Sigma_2)$, where Σ_i is a set of sequences for agent i,

$g = (g_1, g_2)$, where $g_i : \Sigma \to \mathbb{R}$ is the payoff function for agent i,

$\mathcal{K} = (\mathcal{K}_1, \mathcal{K}_2)$, where $\mathcal{K}_i$ is a set of linear constraints on the realization probabilities of agent i.

This model is a two-player game where agent 1 is the defender and agent 2 is the attacker. A *sequence* is an ordered set of player actions taken up to a particular point in the game tree. Let $\mathcal{A}_{i,j}$ be the set of actions available to player i in stage j. Agent 1's fundamental action sequences correspond to allocating the Cyber Mission Forces to physical defense assets. These actions are defined by the set $\mathcal{A}_{1,5}$, where $|\mathcal{A}_{1,5}| = \binom{\tilde{m}_d}{\tilde{n}_d}$. However, to be valid sequences in Σ_1, we index each set of fundamental actions by each information set, so information sets have distinct sequences. Agent 2's fundamental action sequences are $\mathcal{A}_{2,3}$ and $\mathcal{A}_{2,6}$, where $|\mathcal{A}_{2,3}| = \binom{\tilde{m}_a}{\tilde{n}_a}$, corresponding to cyber attacks, and $|\mathcal{A}_{2,6}| = \binom{n_d}{n_a}$, corresponding to physical attacks. Again, Σ_2 indexes each set of fundamental actions by their respective information sets and history of actions.

An *information set* is a set of nodes in the game tree in which a player cannot distinguish between the nodes due to imperfect information about the opponent and nature. Let $\mathcal{I}_i$ be the information sets for player i and $\mathcal{I}_{i,j}$ be the subset of information sets associated with player i and stage j. We have $|\mathcal{I}_{1,5}| = \binom{n_d}{\tilde{m}_d} \sum_{i=0}^{\tilde{n}_a} \binom{n_d}{i}$ for the defender at stage 5, where the first term represents nature's action in stage 1 and the second term represents the defender's information about the attacker's cyber posture in stage 2 and cyber actions in stage 3. Note that even if the cyber sensors are successful in stage 4, the defender cannot distinguish between some combinations of nature's play in stage 2 and the attacker's play in stage 3. Then for the attacker, $|\mathcal{I}_{2,3}| = \binom{n_d}{\tilde{m}_a}$ as the attacker observes nature's action assigning their cyber posture in stage 2, but not the defender's cyber posture in stage 1. Next, $|\mathcal{I}_{2,6}| = |\mathcal{I}_{2,3}| \binom{\tilde{m}_a}{\tilde{n}_a} \binom{n_d}{\tilde{n}_d}$ as the attacker has perfect recall of their cyber posture and cyber attack and we assume they have perfect information on the cyber defense allocation through routine intelligence. However, we assume the attacker does not observe the success of the cyber defense teams prior to launching the physical attack. We assume the attacker

cannot be sure the cyber attack has succeeded until the physical attack because the cyber attack only provides a small window of compromise, as in the United States Air Force Future Operating Concept (2015a).

The overall *payoff* is $g(\sigma) = (g_1(\sigma), g_2(\sigma))$ where $\sigma = (\sigma_1, \sigma_2)$ is pair of sequences for the attacker and defender with $\sigma_1 \in \Sigma_1$ and $\sigma_2 \in \Sigma_2$. All incomplete action sequences have a payoff of zero, that is $g(\sigma) = (0, 0)$ if nature, the defender, and the attacker have not played at every available stage. Let C be the coverage matrix where $c_{i,j}$ is 1 if the air defense located at i covers population center j and 0 otherwise, based on the protection radius, r. Let v_i be the value of city i, defined by its population. Let $a_p(\sigma)$ be the set of indices of the cities targeted by the σ strategy and $a_c(\sigma)$ be the set of indices of the air defense systems targeted by the σ strategy. Similarly, let $d_p(\sigma)$ be the set of indices of the cities in which the air defense systems are located and $d_c(\sigma)$ be the set of indices of the air defense systems defended by the cyber defense teams. Then the payoff function for the attacker is,

$$g_2(\sigma) = \sum_{i \in a_p(\sigma)} v_i \prod_{j \in d_p(\sigma)} (1 - p_{i,j}), \tag{50}$$

where $p_{i,j}$ is the probability city i is protected by air defense asset j. Specifically, cyber and physical defensive assets are successful with independent probabilities such that,

$$p_{i,j} = \begin{cases} 0 & \text{if } C_{i,j} = 0 \\ 0 & \text{if } j \in a_c(\sigma), j \notin d_c(\sigma) \\ p & \text{if } j \notin a_c(\sigma) \\ p\tilde{p}_d(1 - \tilde{p}_a) & \text{if } j \in a_c(\sigma), j \in d_c(\sigma). \end{cases} \tag{51}$$

Notably, the cyber defense and attack effectiveness terms, $\tilde{p}_d(1 - \tilde{p}_a)$, can also be framed as a single cyber defense success parameter. However, by separating the

terms and focusing on effectiveness probabilities, rather than success, we maintain flexibility for modeling asymmetric information, although we maintain information symmetry in this chapter. Since this game is zero-sum, the defender's payoffs are the opposite of the attacker's payoffs, $g_1(\sigma) = -g_2(\sigma)$. Note also that the nature probabilities can be consolidated on leaf nodes by multiplying the leaf value by the probability of reaching that leaf (Koller et al., 1994). This payoff function can be pre-calculated for all sequences and used as a known parameter in the optimization formulations, as long as the size of the joint sequence space fits in memory.

Lastly, $\mathcal{K} = (\mathcal{K}_1, \mathcal{K}_2)$ is the set of linear constraints on the realization probabilities, where $\mathcal{K}_i$ is the set of constraints associated with agent i. We can decompose $\mathcal{K}_1$ into $\mathcal{K}_1 = \mathcal{K}_{1,5}$ where $|\mathcal{K}_{1,5}| = |\mathcal{I}_{1,5}|$. Similarly, we can decompose $\mathcal{K}_2$ into $\mathcal{K}_2 = \mathcal{K}_{2,3} \cup \mathcal{K}_{2,6}$ where $|\mathcal{K}_{2,3}| = |\mathcal{I}_{2,3}|$ and $|\mathcal{K}_{2,6}| = |\mathcal{I}_{2,6}|$. There is one constraint for each information set and each constraint is expressed in terms of $r_i(\sigma_i)$ where $r_i(\sigma_i)$ is a realization plan for a sequence, $\sigma_i \in \Sigma_i$. A *realization plan* is a set of conditional probabilities of selecting each sequence from a given information set, for all information sets in the game. Each constraint enforces the logical requirement that the sum of realization probabilities for all sequences immediately following an information set equals the realization probability of the parent set. Let, $\text{seq}_i : \mathcal{I}_i \to \Sigma_i$ and $\text{Ext}_i : \mathcal{I}_i \to 2^{\Sigma_i}$, as in the notation due to Shoham and Leyton-Brown (2008). The first function, $\text{seq}_i(I)$, returns player i's sequence that led to information set I. The second function, $\text{Ext}_i(I)$, returns all of player i's sequences that extend the sequence that led to $\mathcal{I}_i$ by one action. Then the constraint set, $\mathcal{K}_i$, for player i is,

$$\sum_{\sigma_i' \in \text{Ext}_i(I)} r_i(\sigma_i') = r_i(\text{seq}_i(I)) \quad \forall I \in \mathcal{I}_i. \tag{52}$$

The naive expected utility function for the defender, $g_1(\hat{s}_1, \hat{s}_2)$, given fixed defender

and attacker mixed strategies, $\hat{s}_1$ and $\hat{s}_2$, is

$$\mathbb{E}_{\hat{s}_1}\mathbb{E}_{\hat{s}_2}\left[-\sum_{i\in a_p(\hat{s}_1,\hat{s}_2)} v_i \prod_{j\in d_p(\hat{s}_1,\hat{s}_2)} (1-\phi(\hat{s}_1,\hat{s}_2))\right], \tag{53}$$

where $\phi(\hat{s}_1, \hat{s}_2)$ is a function of integer assignment variables associated with the pure strategies underlying $\hat{s}_1$ and $\hat{s}_2$. Unfortunately, this utility function is a non-linear function of integer variables and, if used directly as an objective function, results in a mixed-integer non-linear program (MINLP) formulation. However, the objective function can be reformulated as a linear utility function by pre-calculating the reward for each discrete combination of player sequences, rather than full mixed strategies. This reformulation, discussed more in Section 6.3, also avoids the need for integer and non-linear constraints, resulting in compact linear programs (LPs) for best-response and Nash equilibrium calculations that scale moderately with problem size. The drawback is that the size of the pre-calculated reward function grows exponentially with problem size, which can limit the scaling of the linear programs.

6.3 Methodology

External sampling linear Monte Carlo counterfactual regret minimization (CFR) is the state-of-the-art for solving large extensive form games in practice (Brown and Sandholm, 2018). However, first-order methods have better theoretical convergence guarantees, $O(\frac{1}{T})$ rather than $O(\frac{1}{\sqrt{T}})$, where T is the number of iterations, and recent results suggest that with appropriate parameter tuning this class of methods can outperform counterfactual regret-based methods (Kroer et al., 2018; Nesterov, 2005). Notably, double oracle column and row generation methods (Bosansky et al., 2014) scale best for problems in which there is an equilibrium with small support (Lisy et al., 2016; Kroer et al., 2018). In this application area, we are interested in mixed

strategies, which have large support over a continuous probability space. Given these prior results, this research focuses on linear programs for exact Nash equilibrium computation and counterfactual regret minimization variants for online, approximate solutions.

6.3.1 Best Response Linear Program

The decision maker can compute a *best response* to the opponent's strategy in polynomial time using a linear program. This linear programming formulation takes advantage of the sequence form representation to avoid an exponential number of decision variables. We accomplish this by directly modeling each player's information sets, action sequences, and realization plans.

As in Section 6.2.2, an information set for player i, $I \in \mathcal{I}_i$, is a set of player i's decision nodes which are indistinguishable due to imperfect information. A sequence for player i, σ_i, in the set of all possible sequences for player i, Σ_i, is an ordered list of all possible combinations of player i's actions, for every level of the game tree. We denote the null action by $\sigma_{i,1}$. A realization plan for player i, $r_i : \Sigma_i \rightarrow \mathbb{R}$, assigns a conditional probability of selecting each sequence, given the agent arrives at the information set leading to that sequence. This formulation supports behavioral mixed strategies which coincide with normal form mixed strategies because our game has perfect recall (Shoham and Leyton-Brown, 2008). In the extensive-form game literature, a *behavioral strategy* is defined by specifying the agent's probability of playing each action at each information set and does not necessarily imply bounded rationality, as is common in other domains. A behavioral mixed strategy randomizes actions at each node in the decision tree, while a traditional mixed strategy randomizes over pure strategies but follows a fixed set of actions within any pure strategy. Recall that $g_i(\sigma_1, \sigma_2)$ is the payoff function for agent i.

Then, following Shoham and Leyton-Brown (2008) with our problem-specific variable and parameter definitions in Section 6.2.2, we have the following linear program for the **Best Response (BR)** problem:

$$\textbf{BR:} \quad \underset{r_1(\sigma_1)\in\mathbb{R}^{|\Sigma_1|}}{\text{maximize}} \quad \sum_{\sigma_1\in\Sigma_1}\left(\sum_{\sigma_2\in\Sigma_2} g_1(\sigma_1,\sigma_2)r_2(\sigma_2)\right) r_1(\sigma_1) \tag{54a}$$

$$\text{subject to} \quad r_1(\sigma_{1,1}) = 1, \tag{54b}$$

$$\sum_{\sigma_1'\in \text{Ext}_1(I)} r_1(\sigma_1') = r_1(\text{seq}_1(I)), \quad \forall\, I \in \mathcal{I}_1, \tag{54c}$$

$$r_1(\sigma_1) \geq 0, \quad \forall\, \sigma_1 \in \Sigma_1. \tag{54d}$$

The objective value of this linear program is the value of the game to agent 1 and the value of r_1 is the behavioral realization plan for agent 1's best response strategy. Note that the set of the defender's best responses to the attacker's equilibrium strategies may include strategies for the defender that are not in the set of the defender's equilibrium strategies. These best response strategies have the same value as the equilibrium strategy, but the attacker's best response to them are not in the attacker's equilibrium set. This happens, for example, in Kuhn poker (Kuhn, 1950).

6.3.2 Nash Equilibrium Linear Program

The agents can also compute *Nash equilibrium* strategies in polynomial time. Consider the dual linear program to the best response linear program. Let v_0 be the value of the game to player 1 and the dual variables $v_I \,\forall\, I \in \mathcal{I}_1$ be the value of the sub-game to player 1 at each information set. Note that in this dual-formulation, $r_2(\sigma_2)$ represents the attacker's behavioral strategy and the dual-value of the dual-variable of v_1 represents the defender's behavioral strategy. We define a convenience function, $\Phi_i : \Sigma_i \to \mathcal{I}_i \cup 0$, that associates each information set with the sequence

that leads to it. For the special null sequence, $\sigma_{i,1}$, the function returns 0.

Then, following Shoham and Leyton-Brown (2008) with our problem-specific variable and parameter definitions in Section 6.2.2, we have the following linear program for the **Nash Equilibrium (NE)** problem:

$$\textbf{NE:} \quad \underset{r_2(\sigma_2)\in\mathbb{R}^{|\Sigma_2|}}{\text{minimize}} \quad v_0 \tag{55a}$$

$$\text{subject to} \quad v_{\Phi_1(\sigma_1)} - \sum_{I'\in\Phi_1(\text{Ext}_1(\sigma_1))} v_{I'} \geq \sum_{\sigma_2\in\Sigma_2} g_1(\sigma_1,\sigma_2) r_2(\sigma_2), \quad \forall\, \sigma_1 \in \Sigma_1, \tag{55b}$$

$$r_2(\sigma_{2,1}) = 1, \tag{55c}$$

$$\sum_{\sigma_2'\in\text{Ext}_2(I)} r_2(\sigma_2') = r_2(\text{seq}_2(I)), \quad \forall\, I \in \mathcal{I}_2, \tag{55d}$$

$$r_2(\sigma_2) \geq 0, \quad \forall\, \sigma_2 \in \Sigma_2. \tag{55e}$$

The objective value of this linear program is the value of the game to agent 1 and the value of r_2 is the behavioral realization plan for agent 2's equilibrium strategy. Agent 1's equilibrium strategy is determined by the dual-value of the v_I variables, which are produced in the solution to the linear program. Note that the Nash equilibrium concept in imperfect-information extensive-form games is weaker than sequential equilibrium, which is an imperfect-information variant of the standard sub-game perfect Nash equilibrium.

6.3.3 Counterfactual Regret Minimization

Counterfactual regret minimization is an approximate technique for solving large extensive-form games (Zinkevich et al., 2008). This algorithm provably converges with a good theoretical and practical convergence rate (Ponsen et al., 2011). CFR has been applied to security games by Lisy et al. (2016) and Davis et al. (2018). Monte Carlo CFR is a chance sampled version of CFR that allows faster iterations for large problems, but takes longer to converge in general (Johanson et al., 2012). There are also a variety of enhancements to the baseline CFR algorithm (Johanson

et al., 2008; Johanson and Bowling, 2009; Brown et al., 2018).

Algorithm 6 shows the basic Monte Carlo CFR algorithm, also referred to as chance sampled CFR. The algorithm includes five primary sub-components: utility evaluation (lines 2-4), Monte Carlo sampling (lines 5-8), regret matching (line 11), utility updating (lines 13-23), and regret updating (lines 24-29). The solve function manages iterations and player updates.

Throughout the algorithm, h is a node that represents the history of player actions, i is the player index, π_i is the accumulated reach probability of player i, and π_{-i} is the accumulated reach probability of player i's opponent. We let ha denote the history determined by h followed by the action a.

In the utility evaluation component, we determine whether the current node, h, is in the set of leaf nodes, $\mathcal{Z}$. Each leaf node has an expected utility based on the chance of reaching that node due to player i, $f_{c,i}(h)$, the chance of reaching that node due to player $-i$, $\pi_{-i}f_{c,-i}(h)$, and the raw utility associated with the node, $u_i(h)$. This utility value is calculated at the leaf nodes of the game tree and, as a result, is performance sensitive.

In the Monte Carlo sampling component, we determine whether the current node, h, is in the set of chance nodes, $\mathcal{C}$. Each chance node has a probability distribution over actions, $f_c(a|h)$, which is used to produce a Monte Carlo sample. In line 9, we use a function to associate each history node with an information set, where each history belongs to exactly one information set. Treating this process as a function rather than a lookup table reduces the memory required and the additional cost of repeated function calls is less than the time penalty for large lookup tables. We use u and u' to track utility and updated utility throughout the algorithm. The regret matching

Algorithm 6 Chance Sampled CFR (Johanson et al., 2012)

function WalkTree(h, i, π_i, π_{-i})
  if $h \in \mathcal{Z}$ **then**
    return $f_{c,i}(h)u_i(h|\pi_{-i}f_{c,-i}(h))$
  end if
  if $h \in \mathcal{C}$ **then**
    Sample outcome $a \in A(h)$ with probability $f_c(a|h)$
    return WalkTree(ha, i, π_i, π_{-i})
  end if
  $I \leftarrow$ information set of h
  $u \leftarrow 0$
  $\sigma \leftarrow$ match regret of I
  for $a \in A(h)$ **do**
    if $player(h) = i$ **then**
      $\pi'_i \leftarrow \sigma[a]\pi_i$
      $u' \leftarrow$ WalkTree(ha, i, π'_i, π_{-i})
      $m[a] \leftarrow u'$
      $u \leftarrow u + \sigma[a]u'$
    else
      $\pi'_{-i} \leftarrow \sigma[a]\pi_{-i}$
      $u' \leftarrow$ WalkTree(ha, i, π_i, π'_{-i})
      $m[a] \leftarrow u'$
      $u \leftarrow u + u'$
    end if
    if $player(h) = i$ **then**
      for $a \in A(I)$ **do**
        $r_I[a] \leftarrow r_I[a] + m[a] - u$
        $s_I[a] \leftarrow s_I[a] + \pi_i\sigma[a]$
      end for
    end if
  end for
  return u
end function

function Solve
  for $t \in \{1, 2, \ldots\}$ **do**
    for $i \in N$ **do**
      WalkTree($\emptyset, i, 1, 1$)
    end for
  end for
end function

component updates the players' current strategies according to scaled regret, where

$$\sigma(I,a) = \begin{cases} \frac{r_I^+[a]}{\sum_{b \in A(I)} r_I^+[b]} & \text{if } \sum_{b \in A(I)} r_I^+[b] > 0 \\ \frac{1}{|A(I)|} & \text{otherwise.} \end{cases} \tag{56}$$

In this regret matching, $r_I[a]$ is the regret for playing action a at information set I. The $+$ operator restricts the matching to positive regret such that $r_I^+[a] = \max\{r_I[a], 0\}$.

In the utility updating component, a is an action in the set of available actions at a given node, $A(h)$, $player(h)$ is the player acting at h, $\sigma[a]$ is the current probability of playing action a, and $m[a]$ holds the updated utility for action a. In this section, we update both the reach probabilities and the current utilities, then recursively call the WalkTree function on the child nodes.

In the regret updating component, at every action a in the set of available actions at a given information set, $A(I)$, we update the regret for the current information set and action, $r_I[a]$, using the counterfactual regret. We also update the cumulative strategy for the current information set and action, $s_I[a]$, using the current strategy and the reach probability. Lastly, the solve function sets the iteration limit and alternates player updates.

In addition to the predetermined iteration limit shown in Algorithm 6, we implement a combined relative-value, time-bound stopping rule parameterized by a relative utility tolerance and a maximum computation time. The relative utility criterion requires that the sample gradient of the defender's estimated utility approach zero with tolerance

$$\delta_n < \epsilon_0 = 5 \times 10^{-5}, \tag{57}$$

where the smoothed sample gradient is calculated as

$$\delta_n = \delta_{n-1}\left(\frac{n-1}{n}\right) + (u_n - u_{n-1})\left(\frac{1}{n}\right). \tag{58}$$

The maximum computation time is set to 10 minutes, unless otherwise noted. When Eq. (57) is met or the computation time is reached, the algorithm terminates and returns the players' utilities and strategies. In our implementation, we treat the time-bound as a soft limit to allow the algorithm to complete sub-tree traversal started before the time limit to avoid partial tree traversals, which results in a small violation of the time-bound on the order of seconds for a 10-minute time-bound.

For moderately-sized problems, we can compare the relative utility tolerance to the exploitability of the defender strategy. The defender utility associated with a given iteration of the CFR algorithm corresponds to the defender's utility given an approximate attacker strategy. However, this is limited in the sense that the attacker's strategy at that iteration is not optimal. To evaluate the true risk of the defender's strategy, we can instead take the defender's strategy from a given iteration and calculate a best attacker response to that strategy. This criterion is more rigorous than the estimated utility, but it requires knowledge of the exact best response strategy.

6.3.4 Discounted Counterfactual Regret Minimization

The default, equally weighted stepsize in the CFR algorithm guarantees convergence and has a fast theoretical convergence rate (Powell, 2007). However, a variant of CFR with adjusted stepsize, CFR+, has been shown to converge more quickly than CFR for a variety of game settings in poker and related domains (Tammelin et al., 2015). Recent results in the CFR family of algorithms have extended the work on stepsize to *discounted CFR* (Brown and Sandholm, 2018). Discounted CFR uses three parameters to control stepsize at iteration t. The α parameter scales accumulated

positive regret by

$$\frac{t^{\alpha}}{t^{\alpha}+1}. \tag{59}$$

The β parameter scales accumulated negative regret by

$$\frac{t^{\beta}}{t^{\beta}+1}. \tag{60}$$

The γ parameter scales contributions to the average strategy by

$$\left(\frac{t}{t+1}\right)^{\gamma}. \tag{61}$$

Brown and Sandholm (2018) find that $\alpha = 1.5, \beta = 0.5, \gamma = 2.0$ are effective parameter settings in the poker domain.

In our Monte Carlo CFR implementation, we consider separate scaling factors for each information set rather than a global iteration factor to take full advantage of the different levels of precision at each information set (Powell, 2007). These changes modify the regret update in Algorithm 6 (line 34-40) such that $r_I[a]$ is scaled using α and β for positive and negative accumulated regret, respectively, and $s_I[a]$ is scaled using γ. The scaling is conducted with respect to each information set's iteration, t_I, rather than the overall iteration, t.

6.4 Testing, Results, and Analysis

To evaluate the exact and approximate approaches, we define a base case scenario with the parameters shown in Table 24 and Table 25. Note that the coverage matrix is defined by the location of the population centers and the coverage radius.

Table 24. Integrated cyber and air defense parameters (base case)

Parameter	Value	Description
$\mathcal{M}$	$\{1, 2, \ldots, 10\}$	Set of population centers
$\mathcal{D}$	$\{1, 3, 4, 5, 8, 9\}$	Set of air defense locations
r	0.3	Coverage radius for each air defense asset
n_d	6	Number of air defense assets
n_a	3	Number of air attack targets
$\tilde{m}_d$	4	Number of cyber defense-capable nodes
$\tilde{m}_a$	4	Number of cyber attack-vulnerable nodes
$\tilde{n}_d$	2	Number of cyber defense teams
$\tilde{n}_a$	2	Number of cyber attack targets
p	0.9	Probability of effective physical defense
$\tilde{p}_s$	0.8	Probability of cyber sensor detection
$\tilde{p}_d$	0.7	Probability of effective cyber defense
$\tilde{p}_a$	0.2	Probability of effective cyber attack

Table 25. Population center location and value (base case, regularized)

City Index	Longitude	Latitude	Value
1	0.03	0.62	0.4797
2	0.74	0.61	0.1596
3	0.80	0.47	0.7116
4	0.11	0.66	0.9690
5	0.37	0.63	0.7139
6	0.64	0.82	0.4747
7	0.99	0.01	0.5066
8	0.85	0.04	0.7074
9	0.29	0.79	0.2100
10	0.98	0.22	0.2952

6.4.1 Results

The solution to the sequence-form linear program for the base case scenario identifies a Nash equilibrium pair of player strategies such that the defender's utility is -1.2595. This linear program has 162,422 decision variables, and 3,347 constraints and solves in 15.49 seconds using the Clp solver (Vesion 1.17) (Lougee-Heimer, 2003) in the Julia programming language (Version 1.0) (Bezanson et al., 2017). In addition to the solve time, the linear program requires the explicit reward function construction and model construction. The reward function is indexed by each player's sequences with 3.21×10^8 total elements, which takes about one hour to construct using a sparse array. The construction of the linear program objective function and constraints using this sparse array is relatively fast at 32.66 seconds. All timing is conducted on a machine with Intel Xeon E5-2680 2.50GHz processors, 24 cores, and 192 GB RAM, and all implementations and results are available at `https://github.com/ajkeith`.

Using the optimal Nash equilibrium strategies and utilities as a benchmark, our results show that the counterfactual regret minimization algorithm finds high quality approximate solutions quickly. Figure 24 shows the raw utility for the defender by iteration of the CFR algorithm. As the number of iterations increases, the algorithm converges to a pair of stationary strategies that approach a Nash equilibrium. The utility is always negative for the defender because there is always a risk the attacker is successful.

Figure 25 shows the performance of the CFR algorithm using *relative exploitability* as a measure of performance. The absolute exploitability, ϵ, of a player's strategy, σ_1, is $\epsilon = u_1(\sigma_1^*, \sigma_2^*) - u_1(\sigma_1, \dot{\sigma}_2)$, where (σ_1^*, σ_2^*) is a Nash equilibrium strategy and $\dot{\sigma}_2$ is a best response to σ_1 (Ponsen et al., 2011). Relative exploitability scales the absolute exploitability to the Nash equilibrium utility and measures the player's expected decrease in utility as a result of departing from a Nash equilibrium, assuming a rational

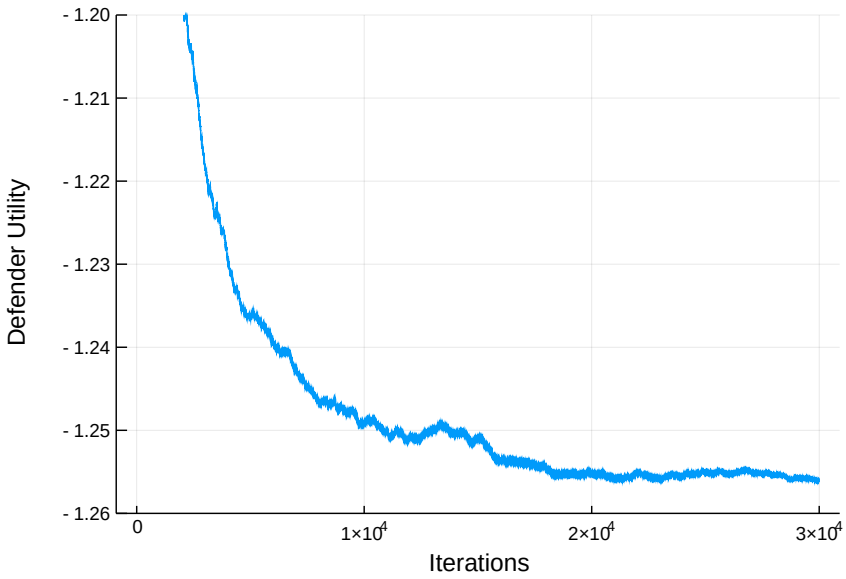

Figure 24. Defender utility (CFR, base case)

adversary. This metric is analogous to an optimality gap in single agent optimization problems. To calculate the relative exploitability of the approximate CFR strategies at a given iteration, we use a sequence-form linear program to determine a best response to the defender's approximate strategy and record the exploitability. This best response linear program has 162,091 variables and 1,366 linear constraints. The results show that the approximate defender strategy produced by CFR converges to within 0.5% relative exploitability of the Nash equilibrium in less than 15,000 iterations, which takes less than 5 minutes at 13 ms per iteration.

The Nash equilibrium strategies produced by these methods are continuous probabilities over the available discrete actions at any given node in the game tree. To illustrate a typical strategy, we select a play sequence that includes the first option at each chance stage and the most likely action for each player at each remaining stage. This six-action sequence corresponds to node $[1, 1, 1, 1, 6, 39]$, where each entry is a numbered action at a game stage. This strategy is depicted visually in Figure 26. In this sub-scenario, both the defender and attacker have the same set of defensive systems that can be affected by their cyber actions: $\{1, 3, 4, 5\}$. The attacker

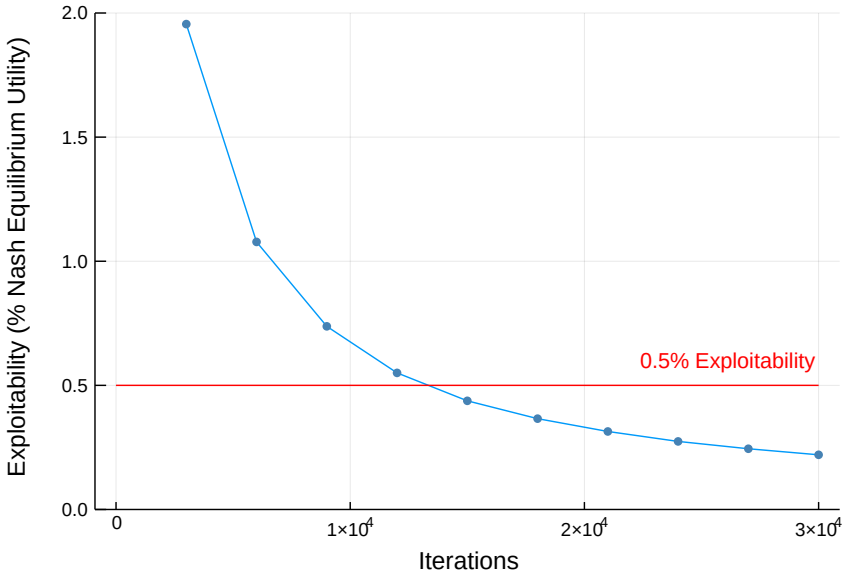

Figure 25. Relative exploitability (CFR, base case)

cyber-attacks air defense asset 3 which has no redundant coverage and air defense asset 1 which has a low likelihood of being defended. Due to chance, the defender's sensors do not detect the cyber attacks so the defender takes a conservative approach of allocating the cyber defense resources to protect air defense assets 4 and 5 because of the high value and coverage. Finally, the attacker physically attacks the naturally undefended city 6, as well as city 2 and city 3 which are undefended due to the cyber attack.

6.4.2 Scaling Properties

To investigate the scaling properties of the linear programming solution approach and the CFR approach, we consider the base case parameters with the total number of cities varying from 7 to 15. We set an upper limit of 12 hours for the calculation of the sequence-form reward functions. We are not able to calculate the full sequence-form reward function for $|\mathcal{M}| = 14$ and $|\mathcal{M}| = 15$ within this time limit, completing only 94% and 70% of the discrete utility values, respectively. The largest game for which we pre-calculated a full sequence-form reward function is $|\mathcal{M}| = 13$ with 9.31×10^6

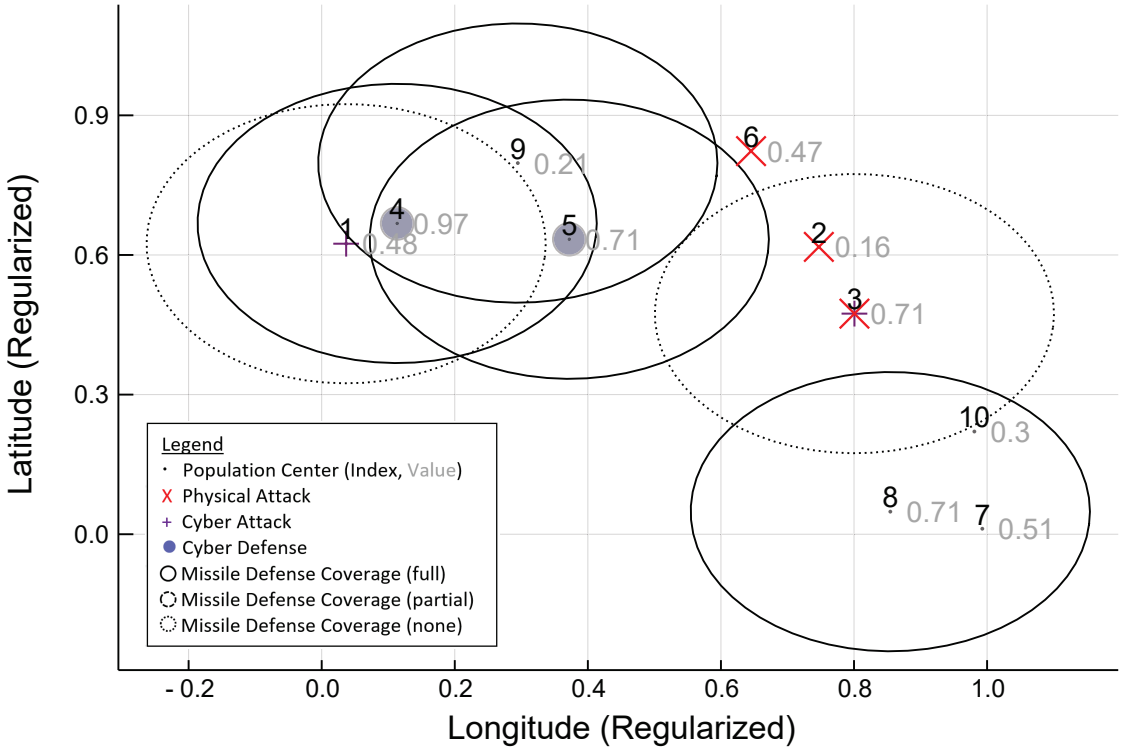

Figure 26. Example action sequence ([1, 1, 1, 1, 6, 39], base case)

history nodes and 7.65×10^8 potential sequence combinations, which completed in 8.38 hours. Because the reward function could not be calculated for the last two parameter settings, we also cannot determine the linear programming solutions for Nash equilibria and best responses.

We set a time limit of 10 minutes and a tolerance of 5×10^{-5} for each CFR solver. Games with $|\mathcal{M}| \geq 13$ were too large to solve to the specified tolerance within the 10 minute time window. However, for $|\mathcal{M}| = 13$, the exploitability is less than 0.5% despite not reaching the tolerance limit. The full results are shown in Table 26. Over the problem instances that can be solved optimally in the time limit, on average, our approximate approach finds solutions in 4.94% of the time it takes to find the Nash equilibrium solution and these approximate solutions have 1.07% more exploitability than the Nash equilibrium solution. Note that the CFR utility column in Table 26 is the defender's utility against an approximate attacker strategy. The

gap column, however, is calculated using the CFR utility of the approximate defender strategy against a best-response attacker, which results in a defender utility that is always lower than or equal to the Nash equilibrium defender utility. The relatively high gaps for $|\mathcal{M}| = 7$ and $|\mathcal{M}| = 10$ are associated with problem instances that have low magnitude exact LP utility. This relationship suggests that a dynamic convergence tolerance which adjusts to the magnitude of the utility might allow for smaller exploitability gaps.

Table 26. Comparison of linear program and CFR solutions by problem size

$\|\mathcal{M}\|$	Sequences ($\times 10^8$)	Reward Time (hrs)	LP Time (sec)	CFR Time (sec)	LP Utility	CFR Utility	Gap (%)
7	0.94	0.07	16	12	-0.684	-0.680	2.71
8	1.50	0.21	25	50	-1.215	-1.218	0.86
9	2.25	0.54	36	258	-1.932	-1.930	0.17
10	3.21	1.21	47	28	-0.500	-0.490	2.74
11	4.41	2.35	64	367	-1.586	-1.576	0.39
12	5.89	4.05	95	572	-1.617	-1.609	0.23
13	7.65	8.38	108	601	-1.655	-1.637	0.37
14	9.74	-	-	601	-	-2.414	-
15	12.2	-	-	601	-	-1.550	-

6.4.3 Parameter Sensitivity

The Nash equilibrium strategy depends on several problem parameters. One area of current operational interest is cyber situational awareness (SA). Detecting and attributing cyber threats can be more challenging than detecting and attributing physical threats. Our model incorporates this practical challenge with the $\tilde{p}_s$ parameter for controlling the defender's likelihood of successful detection for a given cyber attack. The defender's utility is moderately sensitive to this parameter in the base case, as seen in Figure 27. The defender's utility increases as the probability of detection increases, approaching the value of a game with perfect information for SA. The absolute difference is relatively low, indicating that in this case, a decision maker could choose to use a perfect information model with a faster solution time

and approximately a 0.01 decrease in expected utility. Assuming the city value is measured in millions of people, these results can also be used to inform capability investment decisions. For the base case scenario, the sensitivity to probability of detection indicates that improvements in cyber SA translate to a approximately 2,000 lives protected for every 10% increase in probability of cyber detection.

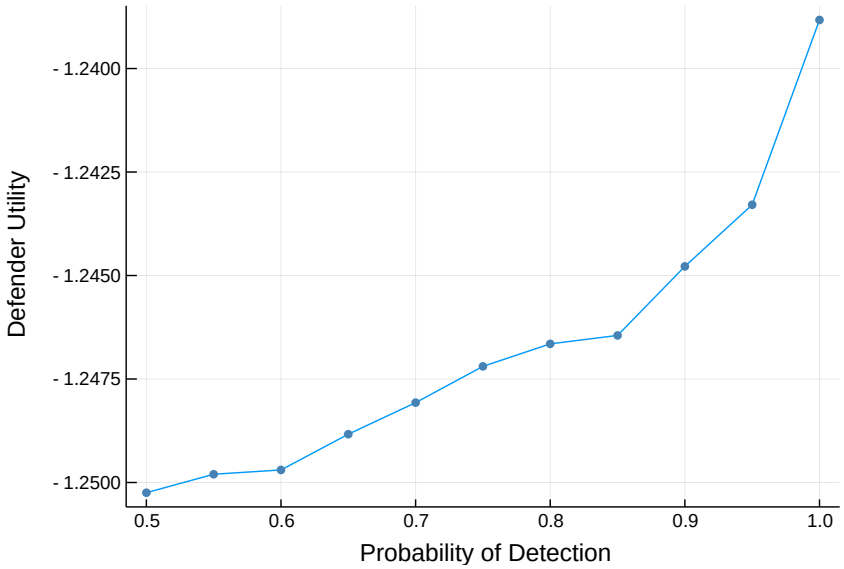

Figure 27. Sensitivity of defender strategy to probability of cyber detection

To investigate the impact of alternative step-sizes for the discounting parameters, we consider standard CFR where $\alpha = \infty, \beta = \infty, \gamma = 0.0$, CFR+ where $\alpha = \infty, \beta = -\infty, \gamma = 2.0$, and discounted CFR where $\alpha = 1.5, \beta = 0.5, \gamma = 2.0$. To approximate the ∞ parameter values, we set the value to 8 which has an error of less than 1×10^{-8} after 10 iterations. We run 50 replications of each of the three parameter combinations and plot the mean and 95% confidence interval of those runs. The results shown in Figure 28 suggest that the discounted CFR parameters lead to the fastest convergence of the three step-size options. On average, discounted CFR with $\alpha = 1.5, \beta = 0.5, \gamma = 2.0$ converges to within 1% error of the Nash equilibrium in 36.4% fewer iterations than standard CFR.

Given these preliminary explorations, we conduct two experiments to investigate

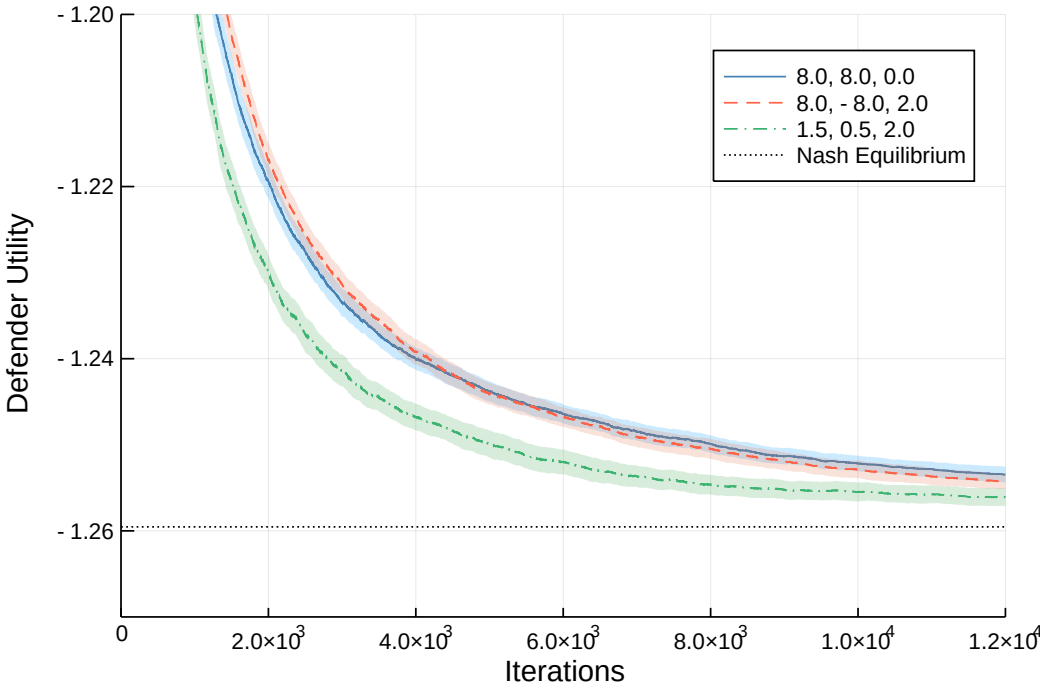

Figure 28. Sensitivity of CFR convergence rate to stepsize parameters (α, β, γ) with mean and 95% confidence interval bands

the impact of problem and algorithm parameters on the strategies produced by the CFR approach. The first experiment is a screening design to identify influential factors. The second design is a space-filling design on a subset of the full parameter space.

In the screening experiment, we construct a 15 factor, resolution IV fractional factorial design with 5 center points and 3 total replications. This design results in 207 runs. The factors and levels are summarized in Table 27. These factors include all factors used to model the problem with modifications to $\tilde{m}_d, \tilde{m}_a, \tilde{n}_d$, and $\tilde{n}_a$ to account for logical restrictions on problem parameter settings (e.g., more defenders than defense-capable nodes). The response is the estimated utility of the defender's strategy. We do not calculate a Nash equilibrium or optimality gap because most problems in the design are too large to be solved with linear programming in a 12-hour time-limit. For the CFR stopping rule, we set a time limit of 30 minutes and a

tolerance of 5×10^{-5}.

Table 27. Experimental factors

Factor	Lower	Upper	Description
$\lvert\mathcal{M}\rvert$	9	15	Number of population centers
r	0.2	0.4	Coverage radius
n_d	4	8	Number of IADS
n_a	2	6	Number of air attack targets
$\tilde{m}_d$ ratio	0.2	1.0	Proportion of cyber defense-capable IADS
$\tilde{m}_a$ ratio	0.2	1.0	Proportion of cyber attack-vulnerable IADS
$\tilde{n}_d$ ratio	0.2	0.8	Ratio of cyber defenses to defense-capable IADS
$\tilde{n}_a$ ratio	0.2	0.8	Ratio of cyber attacks to attack-vulnerable IADS
p	0.6	0.95	Probability of effective physical defense
$\tilde{p}_s$	0.6	0.95	Probability of cyber sensor detection
$\tilde{p}_d$	0.6	0.95	Probability of effective cyber defense
$\tilde{p}_a$	0.05	0.4	Probability of effective cyber attack
α	1	8	Positive regret scale parameter
β	-8	8	Negative regret scale parameter
γ	0	8	Strategy scale parameter

After data collection, we use stepwise regression to fit a least squares regression model to the screening model terms. This analysis produces a model with an adjusted R-squared value above 0.99. The model suggests that all main effects are significant at the 0.005 level along with 45 other interactions. Table 28 shows the main effects, with the $\tilde{m}$ and $\tilde{n}$ factors converted to the original space, rather than the ratio space. To test for over fitting, we fit a model to 70% of the data as a training set and test the model on the remaining 30% as a validation set. The holdback model results in greater than 0.99 R-squared on the validation set. The agreement between the original adjusted R-squared and confirmatory holdback validation analysis indicate the model is not over fitting in a detrimental way, despite the high number of factors in the model. The directional relationship between the response and each factor matches the logic of the scenario. The magnitude of the effects can be grouped into three broad categories: problem size parameters, problem probability parameters, and algorithm parameters. The problem size parameters are the most influential, followed by the problem probability parameters, and finally the algorithm parameters. Although the algorithm tuning parameters are statistically significant, the magnitudes are small.

This indicates that, overall, the effect of the algorithm tuning parameters is not large in this application area.

Table 28. Screening design main effects

Factor	Estimate	P-value
Intercept	-1.391	< 0.0001
$\|\mathcal{M}\|$	-0.148	< 0.0001
r	3.565	< 0.0001
n_d	0.184	< 0.0001
n_a	-0.146	< 0.0001
$\tilde{m}_d$	0.026	< 0.0001
$\tilde{m}_a$	-0.106	< 0.0001
$\tilde{n}_d$	0.139	< 0.0001
$\tilde{n}_a$	-0.183	< 0.0001
p	1.000	< 0.0001
$\tilde{p}_s$	0.057	< 0.0001
$\tilde{p}_d$	0.090	< 0.0001
$\tilde{p}_a$	-0.358	< 0.0001
α	-0.022	< 0.0001
β	0.003	< 0.0001
γ	-0.007	< 0.0001

To investigate the effect of the problem probability parameters and algorithm tuning parameters in more detail, we conduct a second experiment with the problem size fixed at $|\mathcal{M}| = 13, r = 0.3, n_d = 6, n_a = 4, \tilde{m}_d = 6, \tilde{m}_a = 6, \tilde{n}_d = 3, \tilde{n}_a, = 3$. This represents a moderately sized problem with high flexibility for both the defender and attacker, as every IADS node is both defense-capable and attack-vulnerable. We also fix $\tilde{p}_a$ at 0.2 to focus on the defensive probabilities. This game has 2.29×10^6 nodes in the game tree and 2.4×10^8 sequence combinations. Given these fixed settings, we construct a 6-factor, 200-run space-filling design for the remaining factors $(p, \tilde{p}_s, \tilde{p}_d, \alpha, \beta, \gamma)$ with the same limits as in Table 27. The response is the estimated defender utility but we extend the time limit to 120 minutes with the same tolerance of 5×10^{-5}. We use a space-filling design and extend the time limit to provide a more detailed exploration of the probability and parameter tuning space and to identify

any potential non-polynomial effects associated with the tuning parameters.

We use stepwise regression to fit a least squares model to the data from this experiment. This model also fits the data well with an adjusted R-squared above 0.99 and R-squared above 0.99 on a 30% holdback validation set. The main effects for this model are shown in Table 29. No runs in this experiment reach the convergence tolerance within the time limit, indicating there is potential benefit associated with algorithm parameter tuning. However, none of the algorithm tuning parameters are influential for these problem size settings. The β parameter is involved in a second-order interaction with $\tilde{p}_s$ with an estimate of -0.002 and a p-value of < 0.0001. The magnitude of this interaction is small, despite the statistical significance. These results corroborate the findings from the screening model. The algorithm tuning parameters do not have a meaningful effect on solution quality for this specific problem instance or for general problem instances with time limits of 2 hours and 30 minutes, respectively. However, for some problem settings the parameters do affect the convergence rate of the algorithm, as seen in Figure 28. Note that although the time limit is constant in the screening design, the number of iterations varies with problem size. The fewer iterations available for a problem instance, the more parameter tuning affects the utility.

Table 29. Space-filling design main effects

Factor	Estimate	P-value
Intercept	-4.176	< 0.0001
p	1.362	< 0.0001
$\tilde{p}_s$	0.171	< 0.0001
$\tilde{p}_d$	1.351	< 0.0001
β	5.398×10^{-5}	0.2498

6.5 Robust Opponent Exploitation

While the approaches presented in the previous section produce exact and approximate Nash equilibria strategies against rational opponents, these strategies fail to exploit non-equilibria play by opponents with bounded rationality. To address this shortcoming, we consider an extension to exact and approximate robust best responses. A robust best response is a compromise between the maximally conservative Nash equilibria strategy and the maximally aggressive best response strategy.

There are several different approaches to opponent exploitation in the CFR literature, including data biased CFR (Johanson and Bowling, 2009), safe opponent exploitation (Ganzfried and Sandholm, 2015), behaviorally constrained CFR (Farina et al., 2017), Bayesian opponent exploitation (Ganzfried and Sun, 2018), and constrained CFR (Davis et al., 2018). In the following sections, we compare a linear programming formulation for exact robust best response with the data-biased and constrained CFR variants for approximate robust best responses.

6.5.1 Robust Best Response Linear Program

In the opponent exploitation setting, the defender has partial prior knowledge about the opponent's strategy. This knowledge may be derived from play history or subject matter expertise, but is not sufficient to specify an exact opponent strategy. The general problem we need to solve in this case is,

$$\underset{r_1 \in \mathcal{R}_1}{\operatorname{argmax}} \inf_{r_2 \in \mathcal{U}} \sum_{\sigma_1 \in \Sigma_1} \left(\sum_{\sigma_2 \in \Sigma_2} g_1(\sigma_1, \sigma_2) r_2(\sigma_2) \right) r_1(\sigma_1), \tag{62}$$

where $\mathcal{R}_1$ is the set of feasible realization plans for agent 1 and $\mathcal{U}$ is the uncertainty set for the opponent's realization plan. When there is not prior knowledge about the opponent's strategy, the uncertainty set is the entire space of feasible realization

plans and we have the standard best response linear program formulation.

The following formulation shows an explicit robust counterpart to the abstract robust optimization problem using a box uncertainty set. We denote the lower and upper limits of the uncertainty for each component of r_2 by l_{σ_2} and u_{σ_2}, respectively. Note that this formulation is a linear program. However, the constraints associated with (63e) and (63f) add $|\Sigma_2|$ additional constraints to the original linear program, resulting in a substantially larger linear program than the standard best response. We develop the following linear program for the **Robust Best Response (RBR)** problem:

$$\begin{aligned}
\textbf{RBR:} \underset{r_2(\sigma_2)}{\text{minimize}} \quad & v_0 && \text{(63a)}\\
\text{subject to} \quad & v_{\Phi_1(\sigma_1)} - \sum_{I' \in \Phi_1(\text{Ext}_1(\sigma_1))} v_{I'} \geq \sum_{\sigma_2 \in \Sigma_2} g_1(\sigma_1, \sigma_2) r_2(\sigma_2), \forall\, \sigma_1 \in \Sigma_1, && \text{(63b)}\\
& r_2(\sigma_{2,1}) = 1, && \text{(63c)}\\
& \sum_{\sigma_2' \in \text{Ext}_2(I)} r_2(\sigma_2') = r_2(\text{seq}_2(I)), \quad \forall\, I \in \mathcal{I}_2, && \text{(63d)}\\
& r_2(\sigma_2) \geq l_{\sigma_2}, \quad \forall\, \sigma_2 \in \Sigma_2, && \text{(63e)}\\
& r_2(\sigma_2) \leq u_{\sigma_2}, \quad \forall\, \sigma_2 \in \Sigma_2. && \text{(63f)}
\end{aligned}$$

In single decision maker settings, the uncertain parameter is typically a real-valued scalar or vector. However, in game theoretic decision making under uncertainty, the uncertain parameter is the opponent's strategy, which is a probability distribution over actions. As a result, a robust formulation in a game theory setting can also be seen as a distributionally robust formulation, depending on the perspective on the uncertain parameter. In this case, we use a robust framework with uncertainty sets rather than distributionally robust ambiguity sets.

6.5.2 Data-Biased Counterfactual Regret Minimization

In this section, we implement a Monte Carlo variant of data biased CFR (Johanson and Bowling, 2009). In this approach, the opponent (attacker) selects a strategy from

a probabilistically constrained strategy space. The constrained space is defined by a nominal strategy, σ_{fix}, and a confidence parameter for each information set, ρ_I. The confidence parameter controls the likelihood that the opponent plays the nominal strategy at the given information set, such that $\sigma_2(I,a) = \rho_I \sigma_{\text{fix}} + (1-\rho_I)\bar{\sigma_2}(I,a)$ where $\bar{\sigma_2}(I,a)$ is the unconstrained regret-matching strategy. This approach is similar in motivation to ϵ-contamination approaches to uncertainty modeling (Huber, 1964). To calculate the defender's robust strategy, we use Monte Carlo CFR as in Algorithm 6 with the standard regret matching in line 11 replaced with data biased regret matching algorithm such that

$$\sigma_2(I,a) = \begin{cases} \rho_I \sigma_{\text{fix}} + (1-\rho_I)\frac{r_I^+[a]}{\sum_{b \in A(I)} r_I^+[b]} & \text{if } \sum_{b \in A(I)} r_I^+[b] > 0 \\ \rho_I \sigma_{\text{fix}} + (1-\rho_I)\frac{1}{|A(I)|} & \text{otherwise.} \end{cases} \tag{64}$$

Note that only the attacker's regret matching is modified while the defender is able to develop a full best response to the set of possible attack strategies.

6.5.3 Constrained Counterfactual Regret Minimization

Constrained counterfactual regret minimization is an alternative approach to robust CFR developed by Davis et al. (2018). In this approach, the non-equilibrium player is constrained based on their sequence form strategy, rather than their behavioral strategy. Either player can be constrained, but to simplify the notation, we constrain player 1 in the following discussion. The optimization problem in Eq. (62) can be restated by explicitly including the domain constraints in the objective function using Lagrange multipliers,

$$\max_{r_1 \in \mathcal{R}_1} \min_{r_2 \in \mathcal{R}_2, \lambda \geq 0} r_1 G r_2 - \sum_{j=1}^{k} \lambda_j f_j(r_1), \tag{65}$$

where G is the reward matrix associated with $g(\sigma_1, \sigma_2)$, λ is the vector of λ_j Lagrange multipliers for k constraints, denoted by $f_j(r_1)$ (Davis et al., 2018). This Lagrange multiplier approach can be incorporated into CFR by modifying the utility update to penalize constraint violation using the Lagrange multiplier such that

$$\tilde{u} = u - \sum_{j=1}^{k} \lambda_j \nabla_{(I,a)} f_j(\Psi(\sigma_1)), \tag{66}$$

where u is the standard CFR utility, $\tilde{u}$ is the constrained CFR utility, $\nabla_{(I,a)}$ is the gradient with respect to the Ith information set and the ath action, and $\Psi(\sigma_1)$ is the sequence form strategy associated with the behavioral strategy σ_1 (Davis et al., 2018).

In this work, we develop a Monte Carlo variant of constrained CFR, with and without discounting on the positive regret, negative regret, and average strategy. To incorporate the constraint penalty into the Monte Carlo framework, we modify line 21 of Algorithm 6 so that

$$m[a] = u' - \sum_{j=1}^{k} \lambda_j \nabla_{(I,a)} f_j(\Psi(\sigma_1)). \tag{67}$$

Note that constraint penalty is only applied to this action-utility update for the current player at a given node. We do not modify the other utility updates in lines 17, 21, or 22. We also perform gradient descent updates to the Lagrange multipliers prior to the WalkTree call in line 37, such that

$$\lambda_j^t = \max\left(0, \min\left(\lambda_j^{t-1} + \xi^t f_j(\Psi(\sigma_1)), \lambda_{max}\right)\right) \quad \forall j \in \{1, \ldots, k\}, \tag{68}$$

where λ_j^t is the constraint penalty for constraint j on iteration t, ξ^t is the gradient descent learning rate parameter at time t, and λ_{max} is the upper bound on the learn-

ing rate. In the following computational work, we set $\xi^t = \frac{\lambda_{max}^{\lambda_{scale}}}{\sqrt{t}}$ where λ_{scale} is a parameter that scales the learning rate.

6.5.4 Computational Results

In this section, we perform a computational experiment to tune the discounting and constraint parameters for this problem setting, we evaluate the relationship between problem size and computation time, and we compare the standard and robust strategies. We consider the Nash equilibrium strategy, the robust best response LP strategy, the data-biased CFR strategy, and the constrained CFR strategy. Throughout these results, for the nominal attacker strategy, σ_{fix}, we assume an arbitrary pure strategy that assigns probability one to the first action at every information set. To model uncertainty about the opponent's true strategy, we introduce normally distributed noise with mean 0 and standard deviation 0.02. In practice, the nominal strategy and uncertainty set can be derived from previous play history or subject matter expertise.

The Nash equilibrium strategy assumes the opponent will always play a best response to the defender's strategy and, as a result, does not account for irrational strategies like the nominal strategy discussed above. The data biased strategy requires a confidence parameter for each information set. To facilitate comparison with the robust LP approach, we set a constant confidence parameter across information sets at 0.99. For the robust LP approach we define a box uncertainty set based on a 99% confidence interval on the true, normally distributed noise with mean 0 and standard deviation 0.02. This results in a half-width of 0.052 for each information set, with action probabilities saturated at 0 and 1 when the interval would otherwise violate probability limits. We use the same box uncertainty sets as constraints for the constrained CFR model. The 99% information set confidence used in the data-biased

approach and the uncertainty sets based on the 99% confidence interval used in the LP and constrained CFR approaches represent two different ways of modeling the same uncertainty level about the opponent's strategy. Although these uncertainty sets are not equivalent, they are similar from a practical perspective.

To compare these four approaches for opponent modeling, we calculate the Nash equilibrium, robust LP, data-biased CFR, and constrained CFR strategies. Then, we conduct a Monte Carlo simulation that calculates the expected utility for each robust defender strategy against randomly drawn attacker strategies from the distribution of attacker strategies defined by the nominal attacker strategy and the normally distributed noise. Using 180 Monte Carlo samples, we construct empirical distributions and confidence intervals for the mean defender utility under each approach. In addition to the raw utility, we consider the relative exploitability and relative exploitation of each strategy with respect to the Nash equilibrium. By definition, the Nash equilibrium has zero relative exploitability and exploitation. The ideal strategy has high exploitation with low exploitability.

6.5.4.1 Parameter Tuning

This section explores the effect of discount parameters and constraint parameters on the constrained CFR solution. We conduct 3 total replications of a 5-factor, 3-level full factorial experimental design. The factors are the α, β, and γ discount parameters and two constraint parameters, λ_{max} and λ_{scale}, which control the upper bound of the Lagrange penalties and the learning rate of the gradient descent updates for the Lagrange penalties, respectively. Specifically, the λ_{scale} parameter controls the gradient descent learning rate, ξ^t, where $\xi^t = \frac{\lambda_{max}^{\lambda_{scale}}}{\sqrt{t}}$ for iteration t. We have two responses of interest: exploitation and exploitability with respect to the Nash equilibrium strategy. The parameter settings are shown in Table 30. We use the base

case scenario for all robust test cases, with a tolerance of 5×10^{-9} and a 20 minute time limit.

Table 30. Constrained CFR experimental design parameters

	Level		
Factor	Lower	Middle	Upper
α	1	4	8
β	-8	1	8
γ	0	4	8
λ_{max}	10	1000	2000
λ_{scale}	1	4	8

We fit a second-order regression model to the results with exploitation and exploitability as the two responses. The combination of $\lambda_{max} = 2000$ and $\lambda_{scale} = 8$ introduces significant non-polynomial nonlinearities in the response because the strategy fails to converge to a meaningful strategy that outperforms the Nash equilibrium. As a result, we restrict the experimental space to exclude that combination of parameter settings before fitting the second-order model to obtain a better model over the parameter space of interest. To evaluate the effect of removing these data points, we also conduct exploratory analysis including all data points using a second-order model and a neural network to account for the non-linearity. These models confirm the large interaction effect associated with the highest level of these parameters which results in poor performance.

The second-order model for the restricted parameter space results in the coefficients shown in Table 31. The exploitation model has an adjusted R-squared value of 0.96 and the exploitability model has an adjusted R-squared value of 0.84. Note that there are several interactions that are not significant for either response. Removing these factors results in similar adjusted R-squared values, so we keep all factors for completeness. Setting equal weights on exploitation and exploitability and using the second-order model, we find that the following settings maximize exploitation while

Table 31. Second-Order model for constrained CFR tuning

	Exploitation		Exploitability	
Term	Estimate	P-value	Estimate	P-value
Intercept	8.1E-03	$< .0001$	3.3E-02	$< .0001$
α	1.9E-04	$< .0001$	-1.2E-03	$< .0001$
β	2.6E-04	$< .0001$	3.2E-03	$< .0001$
γ	8.8E-04	$< .0001$	-1.9E-03	$< .0001$
λ_{max}	3.5E-06	$< .0001$	2.7E-05	$< .0001$
λ_{scale}	6.3E-05	0.0002	-5.6E-04	0.03
$\alpha * \alpha$	-6.8E-05	$< .0001$	1.8E-05	0.8602
$\alpha * \beta$	-2.9E-05	$< .0001$	-2.7E-04	$< .0001$
$\beta * \beta$	-4.2E-05	$< .0001$	3.1E-04	$< .0001$
$\alpha * \gamma$	-1.0E-06	0.8072	-6.5E-05	0.294
$\beta * \gamma$	-8.5E-06	$< .0001$	-1.6E-06	0.9523
$\gamma * \gamma$	-2.1E-04	$< .0001$	5.1E-04	$< .0001$
$\alpha * \lambda_{max}$	-1.1E-08	0.5437	-1.0E-06	0.0002
$\beta * \lambda_{max}$	-8.9E-08	$< .0001$	2.3E-06	$< .0001$
$\gamma * \lambda_{max}$	4.5E-08	0.0046	-3.6E-07	0.1281
$\lambda_{max} * \lambda_{max}$	-2.8E-09	$< .0001$	-1.8E-08	$< .0001$
$\alpha * \lambda_{scale}$	2.3E-06	0.6538	7.0E-06	0.9284
$\beta * \lambda_{scale}$	-1.4E-06	0.5498	5.8E-05	0.0863
$\gamma * \lambda_{scale}$	-9.3E-08	0.9838	2.2E-04	0.0015
$\lambda_{max} * \lambda_{scale}$	-5.8E-09	0.8197	1.7E-07	0.6502
$\lambda_{scale} * \lambda_{scale}$	-1.0E-05	0.1625	2.3E-04	0.0325

minimizing exploitability,

$$\alpha = 7.4, \beta = -2.9, \gamma = 6.3, \lambda_{max} = 2000, \text{and } \lambda_{scale} = 4.9. \tag{69}$$

6.5.4.2 Computation Time

We also explore the computation time for our implementation of each approach. The results are shown in Table 32 where RLP is the robust linear program, DBCFR is data-biased CFR, and CCFR is constrained CFR. We use a 12 hour limit for the reward function construction and a 30 minute time limit and 25,000 iteration limit for the solvers, to account for the additional complexity of the robust problem. We set a tolerance of 5×10^{-9} for the stopping rule. As in the standard setting, the reward function construction dominates the robust LP solve time. For $|\mathcal{M}| = 14$, the reward function is incomplete at 12 hours, resulting in no reported data for the RLP time and utility for that problem instance.

The robust CFR approaches fail to converge to the specified tolerance for any problem size. However, the robust CFR strategies achieve a defender utility that is close to the robust LP utility while using less than 6% of the LP computation time for $|\mathcal{M}| = 13$. The CFR-based methods produce strategies with defender utilities that are less than 3% degraded with respect to the defender utility of the robust LP, as seen in Figure 29. As such, the CFR methods are appropriate for scaling to large, robust problems.

6.5.4.3 Strategy Comparison

In this section, we compare the exploitation and exploitability of the three robust methods. The 95% confidence interval for the mean defender utility, mean exploitability, and mean exploitation for the base case scenario are shown in Table 33

Table 32. Comparison of robust LP and CFR by problem size

		Time (sec)				Utility		
$\|\mathcal{M}\|$	Seq. ($\times 10^8$)	Reward	RLP	DBCFR	CCFR	RLP	DBCFR	CCFR
7	0.94	252	8	161	710	-0.080	-0.081	-0.082
8	1.50	756	15	158	1098	-0.235	-0.236	-0.236
9	2.25	1944	20	248	1599	-0.579	-0.580	-0.584
10	3.21	4356	39	355	1800	-0.248	-0.249	-0.249
11	4.41	8460	39	530	1800	-0.390	-0.392	-0.395
12	5.89	14580	62	720	1801	-0.605	-0.609	-0.620
13	7.65	30168	80	947	1800	-0.600	-0.607	-0.612
14	9.74	-	-	1294	1802	-	-0.734	-0.743

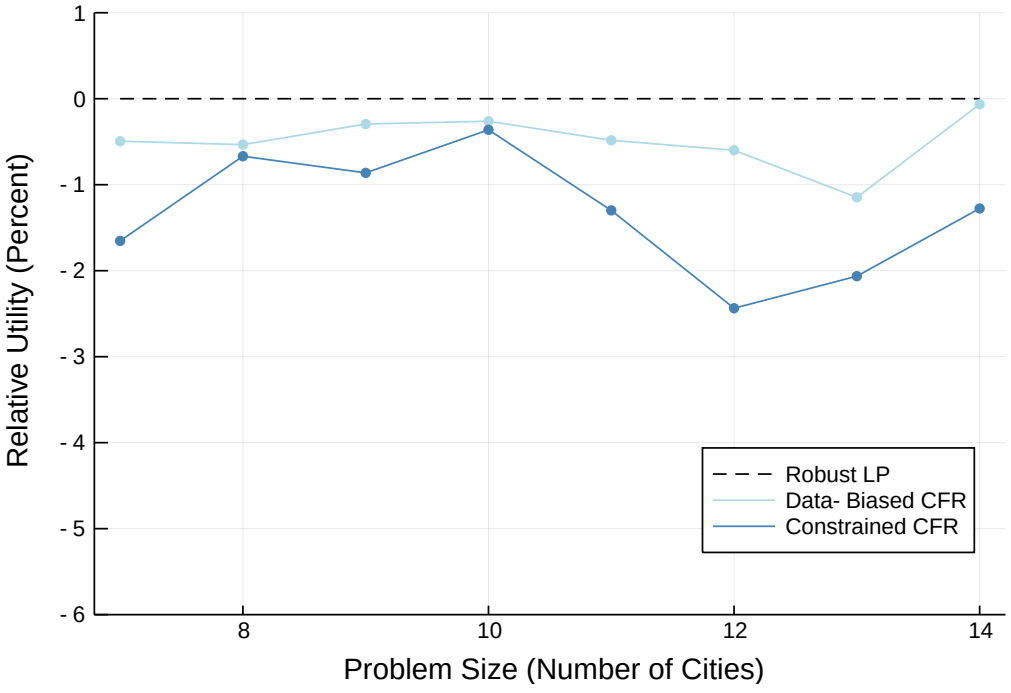

Figure 29. Robust CFR utility relative to robust LP utility

with empirical distributions shown in Figure 30. The robust linear program achieves a favorable exploitation to exploitability ratio of more than 10:1. This strategy allows the decision maker to exploit bounded rationality in an opponent while only risking a fraction of the potential gain if the opponent is in fact perfectly rational. The data-biased CFR approach has a 1:1 ratio of exploitation to exploitability, indicating substantially more risk and lower benefit than the robust LP solution. The constrained CFR approach has approximately a 1:2 ratio of exploitation to exploitability, indicating substantial risk. However, the total exploitation is higher than the data-biased CFR approach. The Nash equilibrium solution is the safest strategy as it guarantees no excess exploitability, but it is also not able to exploit the opponent's mistakes. Overall, the constrained CFR approach results in near-exact solution quality in a small fraction of the computational time.

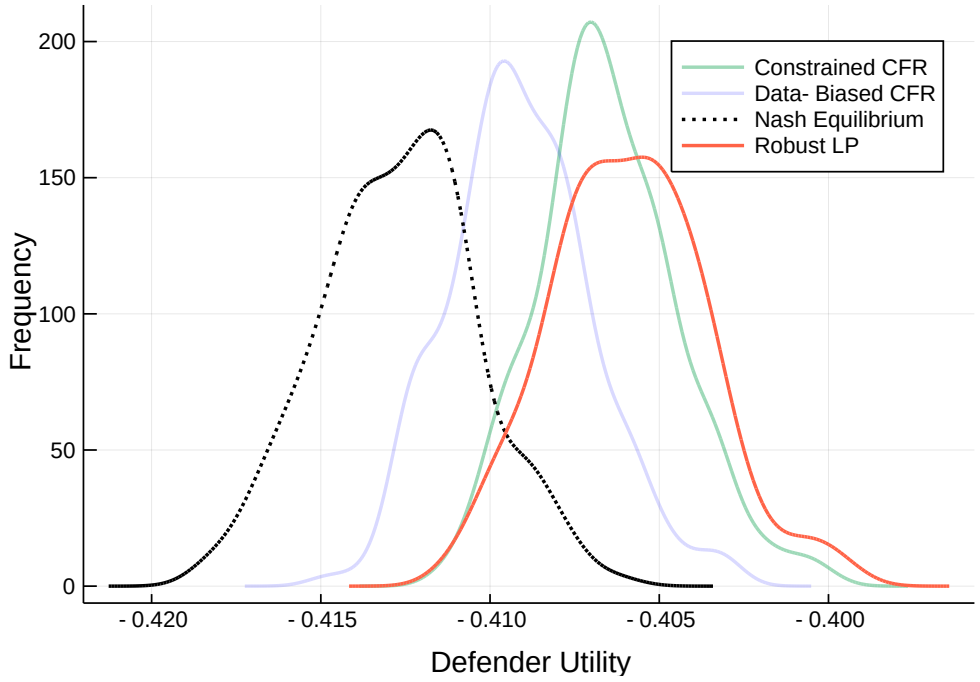

Figure 30. Density of Nash equilibrium and robust strategies against an opponent with bounded rationality (N = 180)

Table 33. Robust strategy performance

Strategy	Mean (95% CI)	Exploitability (%)	Exploitation (%)
Nash Equilibrium	-0.4127 (-0.4131, -0.4125)	0	0
Robust LP	-0.4060 (-0.4063, -0.4058)	0.15	1.63
Data-Biased CFR	-0.4096 (-0.4100, -0.4094)	0.76	0.76
Constrained CFR	-0.4065 (-0.4068, -0.4062)	2.90	1.49

6.6 Conclusion

This research presents multi-domain security games and solves an integrated cyber and air defense problem. This security problem exhibits several challenging features that have not been addressed previously, including imperfect information and mixed strategies for a large, integrated cyber-physical attacker defender game. Realistic operational environments exhibit imperfect information and adaptive adversaries, which are both addressed by this research. Additionally, we present the first implementation of Monte Carlo constrained CFR in the literature and compare it to alternative robust methods.

Our results show that CFR can achieve precise results quickly, with exploitability lower than 0.5% of the optimal exploitability. Furthermore, we quantify the value of cyber situational awareness in this integrated cyber-physical scenario and show the multi-stage impact of improved cyber detection. We also investigate the appropriate parameter settings for the most recent enhancements to traditional CFR by conducting a designed experiment to evaluate the effectiveness of a broad range of stepsizes on algorithm convergence. We also present an extension to robust opponent exploitation. These robust results are the first computational results for Monte Carlo constrained CFR and Monte Carlo constrained CFR with discounting.

This research could be extended in several ways. It would be interesting to directly compare the iterative CFR methods with first-order methods in the domain of attacker-defender games. The excessive gap technique, in particular, would be a useful algorithm for comparative analysis. This game structure is also well suited to

explore opponent exploitation strategies that are based on data-driven methods to statistically learn optimal policies in real time while controlling risk.

This line of research begins to address the emerging challenge of multi-domain operations. It also addresses uncertainty through imperfect information and robust opponent exploitation. As the security environment continues to become more complex, it will be critical to fully integrate cyber and physical security strategies.

VII. Conclusion

This research develops methods and applications for decision making under uncertainty in operations planning, analysis, and assessment. First, a review of the literature is presented to clarify the relationship between modern uncertainty models, decision making models, and optimization models. This theoretically focused review is complemented by a survey of evaluation theory and its applications to military assessment, focusing on practical applications of qualitative uncertainty models. Informed by these literature reviews, the primary methodological and applied research addresses decision making under uncertainty from several perspectives. For static environments, this research develops an order-based server estimation method for G/G/c queues with unobservable parameters. For dynamic environments, this research develops a formulation and solution technique for robust, belief-reward partially observable Markov decision processes (POMDPs). Lastly, to address the most complex case of a dynamic, multi-agent setting, this research presents an application of counterfactual regret minimization (CFR) to an integrated cyber and air defense problem. The key findings of this research clarify the literature, prove theoretical guarantees for new solution techniques, and experimentally evaluate the empirical performance of several new and existing approaches.

7.1 Summary

In Chapter II, the theoretical literature review of decision making and optimization under uncertainty identifies linkages connecting three major areas of research: uncertainty modeling, decision making, and optimization. The optimization literature has used decision-theoretic results for maxmin expected utility to advance the state of the art for optimization under ambiguity. However, the most recent work on

decision theory under ambiguity, including smooth ambiguity and generalized uncertainty averse preferences, has not been incorporated into optimization, suggesting a promising area for further research at the intersection of decision making and optimization. Conversely, some popular uncertainty models, such as fuzzy measures and imprecise probability, have seen several decades of direct implementation but little to no application in decision theory or optimization. This relationship suggests that the modeling capabilities associated with non-set theoretic and non-additive uncertainty models may not justify the additional modeling or computational complexity. The most successful uncertainty, decision, and optimization models address ambiguity in a complete and tractable way, while still maintaining a foundation in probability theory and set theory. This literature review contributes to the field by synthesizing theory and application from three separate research areas to provide new insights to decision making and optimization researchers and practitioners.

Chapter III adopts an applied perspective to survey connections between evaluation theory and the practice of military assessments. The survey highlights the significant overlap between the practice of military assessment and program evaluation. Evaluation theory provides a broad range of assessment paradigms for different contexts, which can be tailored to a particular staff structure or organizational level. Given an assessment paradigm, evaluation and monitoring also provide different levels of rigor and cost for the assessment design and data collection. Assessors can use this suite of options to determine if a descriptive, quasi-experimental, or experimental approach is most appropriate. Then, the assessment report should include appropriate analysis of the qualitative and quantitative uncertainty for the selected design. A mixed-method approach that incorporates both quantitative and qualitative methods produces the highest-quality assessment. This survey contributes to the field by introducing the first classification of military assessment using the dimensions

of evaluation and by providing a guide for practitioners to apply evaluation theory to military assessment.

Chapter IV develops a robust approach for queue inference when the internal parameters of a first-come, first-served (FCFS) G/G/c queue are unknown and the service is non-deterministic. Compared to the existing variance minimization method, the experimental results for a relevant design space show that, in the FCFS setting, the order-based method reaches approximate convergence more quickly, with an 85% reduction in sample size for median convergence, and has similar or lower error before converging, with a 66% reduction in mean estimation error. Theoretical results show that the order-based algorithm has desirable long-run properties in general for the setting of interest as it produces an estimate that is a lower bound and converges to the true value. Furthermore, the order-based algorithm is robust to measurement error in arrival and departure times. These characteristics make the order-based algorithm a suitable estimator for small sample sizes with noisy measurements and limited knowledge of the internal structure of the queue. This contribution includes a new method for robust queue inference, open source code and data, proofs of convergence and lower-bound properties, and experimental results showing improved performance.

Chapter V develops a comprehensive approach to address the challenge of POMDP model misspecification. This approach includes a formulation for robust belief-reward POMDPs. To solve these models, this research extends approximate solution techniques using the new robust belief-reward point-based value iteration algorithm. This algorithm has desirable convexity and convergence properties that make it an efficient approximate solution technique while still addressing the general case of ambiguous dynamics. Experimental results for robust belief-reward POMDPs show that the robust belief-reward value iteration algorithm outperforms standard belief-reward value iteration under worst-case model dynamics for classic problems from the literature.

This worst-case protection is valuable for applications with ambiguous transition and observation dynamics, which are common in practice. Finally, the detailed application of the robust belief-reward formulation and algorithm to a cybersecurity resource allocation problem shows that model misspecification can be particularly problematic in this setting, resulting in an overconfident nominal solution. The robust policy, however, avoids overconfidence and achieves a 61% improvement in the accuracy of intrusion detection compared to the nominal policy. This contribution includes an original formulation for information-collection POMDPs under ambiguity, a new method for solving this class of POMDPs, open source code and data, proofs of convexity and convergence, experimental results showing improved performance, and an application to a cybersecurity problem.

Chapter VI presents an application of counterfactual regret minimization to an integrated cyber and air defense problem. This approach allows the decision maker to address imperfect information in a dynamic environment using mixed strategies. This research presents a reformulation of the original problem into a tractable sequence-form linear program to find an optimal Nash equilibrium solution. Additionally, a new application of the behavioral-form Monte Carlo counterfactual regret minimization algorithm achieves an optimality gap of less than 0.5% in 13,310 iterations which take less than 5 minutes, a 91% reduction in solution time compared to the linear program construction and solution. Empirical results evaluating the value of information for the defensive player show that the defender's expected utility increases with probability of detection at an approximately linear rate in the base scenario. Additionally, this research develops an integrated cyber-air defense application of robust opponent exploitation using linear programming and data biased CFR. This contribution includes a new model that extends security games to integrated cyber and air defense, a new application of counterfactual regret minimization that enables mixed strategies

under imperfect information, open source code and data, and empirical results that show a low optimality gap and a favorable risk-reward trade-off for the robust variant.

7.2 Future Research

This dissertation provides a substantial body of work to be exploited by future research. There are several ways that the methodological and applied work could be extended. For the robust queue inference, there are two primary research directions. First, the current methods could be improved by deriving measures of statistical confidence for each sample size or expected time to converge to a given precision. Second, new methods to solve this problem could provide value by generalizing the context. Approaches that are able to address partially observable customer identity would allow for application to a broader class of problems. It would also be desirable to relax the setting restrictions by extending these results to consider queueing networks rather than simple queues.

The work on robust belief-reward POMDPs is part of an active area of research on robust and ambiguous POMDPs. The primary research direction for this work is extending the current approach to address larger problems. The robust point-based value iteration algorithm could be extended with more sophisticated approximation techniques or adapted to an online implementation. Applying the robustness results from this chapter to large-scale POMDP approximate solution techniques would provide the ability to expand the practical applications by reducing computational time for large-scale problems while still addressing model misspecification. Another possible direction is to explore the policy iteration family of solution techniques. Policy iteration can be more efficient than value iteration for standard POMDPs, which might also be the case for robust POMDPs.

For the adversarial setting, there are two primary directions of interest. First,

other recent advances in opponent exploitation in extensive-form games could be adapted to provide robust solutions for opponents with bounded rationality in the multi-domain security game setting. A comparison of modern robust modeling approaches would help to determine which is most effective in this security setting. Second, there are several variants of counterfactual regret minimization that might provide improved solution times, including public chance sampling and game-tree pruning. In addition to these methodological improvements, it would be interesting to relax some of the scenario assumptions by exploring problems with more cities, higher fidelity cyber attack and defense actions, or multiple rounds of physical attacks.

The literature reviews, methodological advances, and applications in this dissertation improve decision making under uncertainty in operations planning, analysis, and assessment. The contributions provide analysts and decision makers with modeling and algorithmic tools that produce optimal and near-optimal decisions in dynamic, stochastic, ambiguous, and partially observable environments. As the threat environment continues to evolve in unpredictable ways, access to advanced decision making tools will be a key area of strategic advantage.

Appendix A. Robust Queue Inference Proofs and Assumptions

1.1 Proofs

Table 34. Variable definitions

Variable	Definition
A_i	Arrival time of the ith arrival
X_i	Interarrival time following the ith arrival
S_i	Service time of the ith arrival
D_i	Departure time of the ith arrival
$D_{(k,m)}$	kth order statistic among the first m departure times
c	Number of servers
n	Total number of customers
$\tilde{c}_j$	Running estimate for the number of servers following the jth departure
$\hat{c}_j$	Cumulative estimate for the number of servers following the jth departure

Table 35. Assumptions

Index	Assumption
(i)	Customers are uniquely identifiable
(ii)	$\rho < 1$
(iii)	$P(X_i > S_i) > 0$
(iv)	Support of interarrival distribution includes 0
(v)	Interarrival and service distributions are absolutely continuous
(vi)	Interarrival and service distributions are independent

The definitions in Tables 34 and 35 are used throughout the Appendix.

Proposition 2. *Given an FCFS $GI/G/c$ queue that meets Assumptions (i)-(vi), there is a positive probability that the order-based estimation algorithm produces the correct estimate immediately following the cth departure.*

Proof. Let r be the infimum of the support of S_i, t be the supremum of the support of S_i, and s be an arbitrary value in the support of S_i such that $r < s < t$.

We will show by construction that there exist interarrival and service times such that the proposition is true. Let $s_1 = s$. Let $s_i = s - (s - r)\left(\frac{2^i - 1}{2^i}\right)$ for $i \in \{2, 3, \ldots\}$.

Let $x_i \in (0, \frac{s-r}{2^{i+2}})$ for $i \in \{1, 2, \ldots\}$. Using Assumption (v), all s_i are in the support of S_i since for all i, $s_i \in (r, s)$. Using Assumptions (iv) and (v), all x_i are in the support of X_i. For $i = 1$,

$$s_2 + x_1 < s - (s-r)\left(\frac{3}{4}\right) + \frac{s-r}{8} = s - (s-r)\left(\frac{5}{8}\right) < s = s_1. \tag{70}$$

By Assumptions (v) and (vi), this is also true for an arbitrarily small open interval around the particular value of s, so $P(S_2 + X_1 < S_1) > 0$. For $i \in \{2, 3, \ldots, c\}$,

$$\begin{aligned} s_i + x_{i-1} &< s - (s-r)\left(\frac{2^i - 1}{2^i}\right) + \frac{s-r}{2^{i+1}} && (71)\\ &= s - (s-r)\left(\frac{2^{i+1} - 3}{2^{i+1}}\right) && (72)\\ &< s - (s-r)\left(\frac{2^{i-1} - 1}{2^{i-1}}\right) = s_{i-1}. && (73) \end{aligned}$$

By Assumptions (v) and (vi), this is also true for an arbitrarily small open interval around the particular value of s, so $P(S_i + X_{i-1} < S_{i-1} \cap S_{i-1} + X_{i-2} < S_{i-2} \cap \cdots \cap S_2 + X_1 < S_1) > 0$. Then, there is a positive probability that each customer finishes service before the previous customer, which implies a positive probability that the cth customer finishes service before all previous customers. That is,

$$\begin{aligned} P(\hat{c}_c = c) \geq P(&S_c + X_1 + X_2 + \ldots + X_{c-1} < S_1 \cap \\ &S_c + X_1 + X_2 + \ldots + X_{c-1} < S_2 + X_1 \cap \cdots \cap && (74)\\ &S_c + X_1 + X_2 + \ldots + X_{c-1} < S_{c-1} + X_1 + \cdots + X_{c-2}) \\ \geq P(&S_2 + X_1 < S_1 \cap S_3 + X_2 < S_2 \cap \cdots \cap S_c + X_{c-1} < S_{c-1}) && (75)\\ > 0. && (76) \end{aligned}$$

□

Proposition 3. *Given an FCFS $GI/G/c$ queue that meets Assumptions (i)-(vi), the order-based estimation algorithm produces an estimate that is a lower bound on the true number of servers.*

Proof. Let $m \in \{1, 2, \dots\}$ be the number of customers that have departed the queue. Then $\mathcal{D}_m = \{i : D_i \leq D_{(m,m)}\}$ is the set of arrival indices of the first m customers to depart.

Let $\bar{i} = \max \mathcal{D}_m$ be the largest arrival index of customers that have departed. Since this is an FCFS queue, at the time the $\bar{i}$th customer to arrive has departed, all customers with arrival indices $i < \bar{i}$ have already entered service.

Let $\mathcal{S}_m = \{1, 2, \dots, \bar{i}\} / \mathcal{D}_m$ be the set of arrival indices of customers known to have entered service but not yet departed, immediately after the mth departure. Then $\tilde{c}_m = |\mathcal{S}_m| + 1$. Immediately prior to the mth departure, there are at least $\tilde{c}_m$ customers in service: the mth customer to depart and all the customers whose arrival indices are in $\mathcal{S}_m$. Thus, there are at least $\tilde{c}_m$ servers and $\hat{c}_m \leq c$. □

Using Proposition 2, Theorem 4 shows that the order-based estimator converges in probability to the true number of servers. The argument of the proof is that for any observation period in which the systems starts empty, there is a positive probability the cth customer to arrive will be the first customer to leave service, which results in a correct running and cumulative estimate. Furthermore, busy cycles occur infinitely often with probability one. As the number of observed customers approaches infinity, the probability of no correct running estimates decays to zero and the probability of a correct cumulative estimate approaches one.

Theorem 4. *For an FCFS $GI/G/c$ queue that meets Assumptions (i)-(vi), the order-based server estimation algorithm produces estimates $\hat{c}_n$ such that* $\lim_{n\to\infty} P(|\hat{c}_n - c| > \epsilon) = 0 \quad \forall\, \epsilon > 0$.

Proof. Let $\epsilon_1 \geq 1$, then $P(|\hat{c}_n - c| > \epsilon_1) \leq P(|\hat{c}_n - c| > \epsilon)$ as $[\epsilon_1, \infty) \subseteq (\epsilon, \infty)$. Without loss of generality, let $0 < \epsilon < 1$. As $c, \hat{c}_n \in \mathbb{Z}^+$, $P(|\hat{c}_n - c| > \epsilon) = P(\hat{c}_n \neq c)$.

Let n be the total number of customers observed, $b(n)$ be the number of busy periods, and m_k be the number of customers in busy period k. Note that by Proposition 2, for any busy period there is a positive probability that $m_k \geq c$ and that given $m_k \geq c$, $P(\hat{c}_{m_k} = c) \geq P(\hat{c}_c = c) > 0$. Then,

$$P(\hat{c}_n \neq c) = \prod_{k=1}^{b(n)} P(\hat{c}_{m_k} \neq c) \leq P(\hat{c}_c \neq c)^{b(n)} \qquad \text{where } P(\hat{c}_c \neq c) < 1. \tag{77}$$

By Assumptions (ii) and (iii), the system will be empty infinitely often with probability one (Whitt, 1972, Theorem 2.2), so $\lim_{n\to\infty} b(n) = \infty$. Then,

$$\lim_{n\to\infty} P(|\hat{c}_n - c| > \epsilon) = \lim_{n\to\infty} P(\hat{c}_n \neq c) \leq \lim_{n\to\infty} P(\hat{c}_c \neq c)^{b(n)} = 0 \tag{78}$$

$$\Rightarrow \lim_{n\to\infty} P(|\hat{c}_n - c| > \epsilon) = 0. \tag{79}$$

□

We use the same proof structure from Proposition 2, Proposition 3, and Theorem 4 to show that the LCFS order-based estimator produces an estimate which is a lower bound on the true number of servers and that the LCFS order-based estimator converges in probability to the true number of servers. The key argument in the proof for Proposition 5 is that a particular relationship between arrivals and departures has positive probability and that the order-based estimate is correct given that relationship. There are two key arguments in the proof for Proposition 6. First, we note that all customers that were the last to arrive at the time of another customer's departure (i.e., highest priority in the LCFS queue at that time) and who have not departed must be in service simultaneously. Second, we show that the order-based estimate

produces the same number.

Proposition 5. *Given an LCFS GI/G/c queue that meets Assumptions (i)-(vi), there is a positive probability that the LCFS order-based estimation algorithm produces the correct estimate immediately following the cth departure.*

Proof. Let $\mathcal{S}_t$ be the set of arrival indices of customers known to have entered service by time t. By Assumptions (iv), (v), and (vi), there is positive joint probability of a positive service time and arbitrarily small interarrival time for each customer. There is positive probability that the first $2c$ customers arrive before any departures, so $P(X_1 + X_2 + \cdots + X_{2c-1} < \min\{S_1, S_2, \ldots, S_c\}) > 0$.

Let $\mathcal{I}_{c+1,2c}$ be the arrival indices of the departures $c+1, c+2, \ldots, 2c$. Given the first $2c$ customers arrive before any departures, by Assumptions (v) and (vi), there is positive joint probability that the first c customers depart before any of the customers with higher arrival indices, so $P(\cap_{i \in \mathcal{I}_{c+1,2c}} S_i > \max\{S_1, S_2, \ldots, S_c\} \mid X_1 + X_2 + \cdots + X_{2c-1} < \min\{S_1, S_2, \ldots, S_c\}) > 0$. For each of these c departures, the server immediately begins serving the next highest priority customer in the LCFS queue since the queue is non-empty with positive probability.

Let t_c be the time of the cth departure. Then there is positive probability that $\mathcal{S}_{t_c} = \mathcal{I}_{c+1,2c}$ and $|\mathcal{S}_{t_c}| = c$. Thus, $P(\hat{c}_c = c) > 0$. □

Proposition 6. *Given an LCFS GI/G/c queue that meets Assumptions (i)-(vi), the LCFS order-based estimation algorithm produces an estimate that is a lower bound on the true number of servers.*

Proof. Let $m \in \{1, 2, \ldots\}$ be the number of customers that have departed the queue. For $i \in \{1, 2, \ldots\}$, let $\mathcal{A}_t = \{i : A_i \leq t\}$ be the set of arrival indices of the customers that arrive by time t and let $\mathcal{D}_t = \{i : D_i \leq t\}$ be the set of arrival indices of the customers that depart by time t, where D_i is the departure time of the ith arrival.

Let $\mathcal{Y}_t = \mathcal{A}_t \,/\, \mathcal{D}_t$ be the set of arrival indices of customers in queue or service at time t. Let $\mathcal{S}_t$ be the set of arrival indices of customers known to have entered service at time t. Since the queue is LCFS, let $\mathcal{S}_{D_i} = (\mathcal{S}_{D_i-\epsilon} \,/\, \{i\}) \cup \max \mathcal{Y}_{D_i}$ where $\epsilon > 0$ is arbitrarily small and let $\mathcal{S}_t = \varnothing \,\forall\, t < D_1$. That is, when a customer departs, that departing customer is removed from the set of customers known to be in service and the highest priority customer in the system is added to the set of customers known to be in service.

Then $\tilde{c}_m = |\mathcal{S}_{D_m}|$. Immediately following the mth departure, there are at least $\tilde{c}_m$ customers in service: all customers that were the last to arrive at the time of another customer's departure and who have not departed yet. Thus, there are at least $\tilde{c}_m$ servers and $\hat{c}_m \leq c$. □

Theorem 7. *For an LCFS GI/G/c queue that meets Assumptions (i)-(vi), the LCFS order-based server estimation algorithm produces estimates $\hat{c}_n$ such that* $\lim_{n\to\infty} P(|\hat{c}_n - c| > \epsilon) = 0 \quad \forall\, \epsilon > 0$.

Proof. Let $\epsilon_1 \geq 1$, then $P(|\hat{c}_n - c| > \epsilon_1) \leq P(|\hat{c}_n - c| > \epsilon)$ as $[\epsilon_1, \infty) \subseteq (\epsilon, \infty)$. Without loss of generality, let $0 < \epsilon < 1$. As $c, \hat{c}_n \in \mathbb{Z}^+$, $P(|\hat{c}_n - c| > \epsilon) = P(\hat{c}_n \neq c)$.

Let n be the total number of customers observed, $b(n)$ be the number of busy periods, and m_k be the number of customers in busy period k. Note that by Proposition 5, for any busy period there is a positive probability that $m_k \geq 2c$ and that given $m_k \geq 2c$, $P(\hat{c}_{m_k} = c) \geq P(\hat{c}_c = c) > 0$. Then,

$$P(\hat{c}_n \neq c) = \prod_{k=1}^{b(n)} P(\hat{c}_{m_k} \neq c) \leq P(\hat{c}_c \neq c)^{b(n)} \qquad \text{where } P(\hat{c}_c \neq c) < 1. \tag{80}$$

By Assumptions (ii) and (iii), the system will be empty infinitely often with prob-

ability one (Whitt, 1972, Theorem 2.2), so $\lim_{n\to\infty} b(n) = \infty$. Then,

$$\lim_{n\to\infty} P(|\hat{c}_n - c| > \epsilon) = \lim_{n\to\infty} P(\hat{c}_n \neq c) \leq \lim_{n\to\infty} P(\hat{c}_c \neq c)^{b(n)} = 0 \tag{81}$$

$$\Rightarrow \lim_{n\to\infty} P(|\hat{c}_n - c| > \epsilon) = 0. \tag{82}$$

□

1.2 Assumptions

This section discusses the motivation for each assumption in Table 35 and presents counterexamples to highlight the impact of Assumptions (iv) - (vi).

Assumption (i).

Assumption (i) is that customers are uniquely identifiable. Assumption (i) is required for Proposition 2 to hold. Without Assumption (i), the order-based estimate is not well-defined.

Assumption (ii).

Assumption (ii) is that $\rho < 1$. Assumption (ii) is required for Theorem 4 to hold (Whitt, 1972, Theorem 2.2).

Assumption (iii).

Assumption (iii) is that $P(X_i > S_i) > 0$. Assumption (iii) is required for Theorem 4 to hold (Whitt, 1972, Theorem 2.2).

Assumption (iv).

Assumption (iv) is that the support of the interarrival distribution includes 0. Given only Assumptions (i), (ii), (iii), (v), and (vi), the following counterexample shows that Proposition 2 does not hold in the absence of Assumption (iv).

Suppose customers are uniquely identifiable, X_i and S_i are independent, $X_i \sim \text{Unif}(0.5, 1), S_i \sim \text{Unif}(0, 0.9)$, and $c = 3$, where X_i is the interarrival time following the ith arrival, S_i is the service time of the ith arrival, and c is the number of servers. Then we have,

Assumption (i): True, by assumption.

Assumption (ii): True, as $E[X_i] = 0.75, E[S_i] = 0.45 \Rightarrow \rho < 1$.

Assumption (iii): True, as $P(X_i > 0.8 \cap S_i < 0.8) > 0 \Rightarrow P(X_i > S_i) > 0$.

Assumption (iv): False, as $0 \notin \text{Support}(X_i)$.

Assumption (v): True, as X_i and S_i follow Uniform distributions.

Assumption (vi): True, by assumption.

The minimum possible departure time for the third customer is 1.0 and the maximum possible departure time of the first customer is 0.9. So, the first customer always departs before the third customer arrives and the order-based estimator cannot produce the correct estimate in the first c customers. This counterexample supports the necessity of Assumption (iv).

Assumption (v).

Assumption (v) is that the interarrival and service distributions are absolutely continuous and therefore non-deterministic. Given only Assumptions (i), (ii), (iii),

(iv), and (vi), the following counterexample shows that Proposition 2 does not hold in the absence of Assumption (v).

Suppose customers are uniquely identifiable, X_i and S_i are independent, $X_i \sim \text{Unif}(0, 2)$, $S_i \sim \text{Deterministic}(0.9)$, and $c = 3$. Then we have,

Assumption (i): True, by assumption.

Assumption (ii): True, as $E[X_i] = 1, E[S_i] = 0.9 \Rightarrow \rho < 1$.

Assumption (iii): True, as $P(X_i > 1 \cap S_i < 1) > 0 \Rightarrow P(X_i > S_i) > 0$.

Assumption (iv): True, as $0 \in \text{Support}(X_i)$.

Assumption (v): False, as the distribution of S_i is a deterministic point mass.

Assumption (vi): True, by assumption.

As the service time is deterministic, customers will depart in the same order in which they arrived, regardless of interarrival time. This guarantees an incorrect order-based estimate after three departures. This counterexample supports the necessity of Assumption (v).

Assumption (vi).

Assumption (vi) is that the interarrival and service distributions are independent. Given only Assumptions (i), (ii), (iii), (iv), and (v), the following counterexample shows that Proposition 2 does not hold in the absence of Assumption (vi).

Suppose customers are uniquely identifiable,

$$X_i \sim \text{Unif}(0, 1) \tag{83}$$

$$T \sim \text{Triangular}(0.25, 0.75, 0.25) \tag{84}$$

$$S_i = \begin{cases} \frac{1}{2}X_i & \text{if } X_i < 0.5 \\ T & \text{otherwise,} \end{cases} \tag{85}$$

and $c = 3$, where T is a dummy triangular distribution. Then we have,

Assumption (i): True, by assumption.

Assumption (ii): True, as $E[X_i] = 0.5, E[S_i] = 0.2708 \Rightarrow \rho < 1$.

Assumption (iii): True, as $P(X_i > 0.6 \cap S_i < 0.6) > 0 \Rightarrow P(X_i > S_i) > 0$.

Assumption (iv): True, as $0 \in \text{Support}(X_i)$.

Assumption (v): True, as the distribution of X_i is Uniform, and the distribution of S_i is absolutely continuous over $[0, 0.75]$ including at 0.25.

Assumption (vi): False, as S_i is a function of X_i.

If either interarrival time is less than 0.5, the earlier customer departs before the arrival of the next customer and the order-based estimator cannot produce the correct estimate in the first c customers. Otherwise, if both interarrival times are greater than or equal to 0.5, the minimum possible departure time for the third customer is 1.0. The maximum possible departure time of the first customer is 0.75. So, the first or second customer always departs before the arrival of the third customer and the order-based estimator cannot produce the correct estimate in the first c customers. This counterexample supports the necessity of Assumption (vi).

1.3 Code

Full code with documentation and testing is available at `https://github.com/ajkeith/UnobservableQueue.jl`. Experimental simulation data and benchmarks are also available at the same location.

Appendix B. Robust POMDP Proofs and Problem Formulations

2.1 Proofs

Proof of Theorem 8.

The approach of the proof for Theorem 8 is to use Theorem 3.1 of Araya-López et al. (2010) to extend Theorem 1 of Osogami (2015) to the belief-reward setting. The structure of the proof for 8 is step-wise analogous to the structure of the proof for Theorem 1 in the research conducted by Osogami (2015), with variations to address the robust belief-reward aspect of this work.

Theorem 8. *When $N < \infty$, the robust belief-reward value function, $V_n(\mathbf{b})$, is convex with respect to $\mathbf{b}$ for each $n \in [0, N]$ for a convex ambiguity set, $\mathcal{P}_s^a$ for $s, a \in \mathcal{S} \times \mathcal{A}$, and a convex belief-reward function, $\rho(\mathbf{b}, a)$.*

Proof. Proof. For $n \in [0, N)$, the robust belief-reward Bellman equation follows directly from Osogami (2015, Theorem 1) and Araya-López et al. (2010, Theorem 3.1):

$$V_n(\mathbf{b}) = \max_{a \in \mathcal{A}} \rho(\mathbf{b}, a) + \min_{p_n^a \in \mathcal{P}^a} \sum_{s \in \mathcal{S}} \left(\mathbf{b}(s) \gamma \sum_{t,z \in \mathcal{S} \times \mathcal{Z}} p_n^a(t, z|s) V_{n+1}(\mathbf{b'}_{\mathbf{b},a,z}) \right), \tag{86}$$

where $\mathbf{b'}_{\mathbf{b},a,z}$ is the belief state after taking action $a \in \mathcal{A}$ from $\mathbf{b} \in \mathcal{B}$ and observing $z \in \mathcal{Z}$:

$$\mathbf{b}'_{\mathbf{b},a,z}(t) = \frac{\sum_{s \in \mathcal{S}} p_n^a(t, z|s) \mathbf{b}(s)}{\sum_{s',t' \in \mathcal{S}^2} p_n^a(t', z|s') \mathbf{b}(s')}, \forall t \in \mathcal{S}. \tag{87}$$

By the induction hypothesis in Osogami (2015, Theorem 1), there exists a possibly infinite set of vectors, Λ_{n+1}, such that

$$V_{n+1}(\mathbf{b}) = \max_{\alpha \in \Lambda_{n+1}} \left[\sum_{s \in \mathcal{S}} \alpha(s) \mathbf{b}(s) \right]. \tag{88}$$

Substituting (88) into (86), we have

$$V_n(\mathbf{b}) = \max_{a\in\mathcal{A}} \rho(\mathbf{b},a) + \min_{p_n^a\in\mathcal{P}^a} \sum_{s\in\mathcal{S}} \left(\mathbf{b}(s)\gamma \sum_{t,z\in\mathcal{S}\times\mathcal{Z}} p_n^a(t,z|s) \max_{\alpha_z\in\Lambda_{n+1}^{(a,z)}} \sum_{x\in\mathcal{S}} \mathbf{b'}_{\mathbf{b},a,z}(x)\alpha_z(x) \right). \tag{89}$$

Substituting (87) into (89), we have

$$V_n(\mathbf{b}) = \max_{a\in\mathcal{A}} \rho(\mathbf{b},a) + \min_{p_n^a\in\mathcal{P}^a} \left(\gamma \sum_{z\in\mathcal{Z}} \max_{\alpha_z\in\Lambda_{n+1}^{(a,z)}} \sum_{x,s'\in\mathcal{S}^2} p_n^a(x,z|s')\mathbf{b'}(s')\alpha_z(x) \right) \tag{90}$$

$$= \max_{a\in\mathcal{A}} \rho(\mathbf{b},a) + \min_{p_n^a\in\mathcal{P}^a} \sum_{z\in\mathcal{Z}} \max_{\alpha_z\in\Lambda_{n+1}^{(a,z)}} \sum_{s\in\mathcal{S}} \mathbf{b}(s) \left(\gamma \sum_{t\in\mathcal{S}} p_n^a(t,z|s)\alpha_z(t) \right), \tag{91}$$

where the variables s' and x are changed to s and t, respectively, in the last equality.

For a given $p_n^a(\cdot,z|s)$ and a given α_z,

$$\sum_{s\in\mathcal{S}} \mathbf{b}(s) \left(\gamma \sum_{t\in\mathcal{S}} p_n^a(t,z|s)\alpha_z(t) \right) \tag{92}$$

is a linear function of $\mathbf{b}$. The maximum of linear functions is convex and the sum of convex functions is convex, so

$$\sum_{z\in\mathcal{Z}} \max_{\alpha_z\in\Lambda_{n+1}^{(a,z)}} \sum_{s\in\mathcal{S}} \mathbf{b}(s) \left(\gamma \sum_{t\in\mathcal{S}} p_n^a(t,z|s)\alpha_z(t) \right) \tag{93}$$

is a convex function of $\mathbf{b}$. Let $Q_n^a(\mathbf{b})$ be defined as follows,

$$Q_n^a(\mathbf{b}) \equiv \min_{p_n^a\in\mathcal{P}^a} \sum_{z\in\mathcal{Z}} \max_{\alpha_z\in\Lambda_{n+1}^{(a,z)}} \sum_{s\in\mathcal{S}} \mathbf{b}(s) \left(\gamma \sum_{t\in\mathcal{S}} p_n^a(t,z|s)\alpha_z(t) \right). \tag{94}$$

Exchange the summation over z and maximum over α_z in (94) to obtain

$$Q_n^a(\mathbf{b}) \equiv \min_{p_n^a\in\mathcal{P}^a} \max_{\alpha_z\in\Lambda_{n+1}^{(a,z)}} \sum_{z\in\mathcal{Z}} \sum_{s\in\mathcal{S}} \mathbf{b}(s) \left(\gamma \sum_{t\in\mathcal{S}} p_n^a(t,z|s)\alpha_z(t) \right), \tag{95}$$

and define M as

$$M \equiv \sum_{z\in\mathcal{Z}} \sum_{s\in\mathcal{S}} \mathbf{b}(s) \left(\gamma \sum_{t\in\mathcal{S}} p_n^a(t, z|s)\alpha_z(t) \right). \tag{96}$$

This M is the same as the M from Osogami (2015, Theorem 1) with the exception of the removal of the reward term, which is constant with respect to p_n^a and α_z. Thus, Loomis' Minimax Theorem in Motwani and Raghavan (1995, Theorem 2.3) remains valid and we have

$$\min_{p_n^a\in\mathcal{P}^a} \max_{\alpha_z\in\Lambda_{n+1}^{(a,z)},z\in\mathcal{Z}} M = \max_{\alpha_z\in\bar{\Lambda}_{n+1}^{(a,z)},z\in\mathcal{Z}} \min_{p_n^a\in\mathcal{P}^a} M, \tag{97}$$

where $\bar{\Lambda}_{n+1}^{(a,z)}$ is the convex hull of $\Lambda_{n+1}^{(a,z)}$. By (97) and (94),

$$Q_n^a(\mathbf{b}) = \max_{\alpha_z\in\bar{\Lambda}_{n+1}^{(a,z)},z\in\mathcal{Z}} \min_{p_n^a\in\mathcal{P}^a} \sum_{z\in\mathcal{Z}} \sum_{s\in\mathcal{S}} \mathbf{b}(s) \left(\gamma \sum_{t\in\mathcal{S}} p_n^a(t, z|s)\alpha_z(t) \right) \tag{98}$$

$$= \max_{\alpha_z\in\bar{\Lambda}_{n+1}^{(a,z)},z\in\mathcal{Z}} \sum_{s\in\mathcal{S}} \mathbf{b}(s) \min_{p_n^a(\cdot,\cdot|s)\in\mathcal{P}_s^a} \left(\gamma \sum_{t,z\in\mathcal{S}\times\mathcal{Z}} p_n^a(t, z|s)\alpha_z(t) \right). \tag{99}$$

Thus $Q_n^a(\mathbf{b})$ can be represented by the maximum of a possibly infinite number of functions that are linear with respect to $\mathbf{b}$ and $Q_n^a(\mathbf{b})$ is convex with respect to $\mathbf{b}$. By assumption $\rho(\mathbf{b}, a)$ is also convex. Then, because the sum of convex functions is convex and the maximum of convex functions is convex, we have that

$$V_n(\mathbf{b}) = \max_{a\in\mathcal{A}} \rho(\mathbf{b}, a) + Q_n^a(\mathbf{b}) \tag{100}$$

is convex when $\mathcal{P}_s^a$ and $\rho(\mathbf{b}, a)$ are convex. □

Proof of Theorem 9.

The approach of the proof for Theorem 9 is to use Theorem 3.1 of Araya-López et al. (2010) to extend Theorem 2 of Osogami (2015) to the belief-reward setting. The

structure of the proof for 9 is step-wise analogous to the structure of the proof for Theorem 2 in the research conducted by Osogami (2015), with variations to address the robust belief-reward aspect of this work.

Theorem 9. *The robust belief reward value function, V_0, satisfying $V_n(\mathbf{b}) = \max_{a\in\mathcal{A}} \rho(\mathbf{b},a)+\min_{p_n^a\in\mathcal{P}^a} \sum_{s\in\mathcal{S}} \left(\mathbf{b}(s)\gamma \sum_{t,z\in\mathcal{S}\times\mathcal{Z}} p_n^a(t,z|s)V_{n+1}(\mathbf{b'}_{\mathbf{b},a,z})\right)$, converges uniformly as $N\to\infty$ if $\gamma<1$.*

Proof. Proof. Let $\mathcal{L}$ be the operator that maps $V_{n+1}(\cdot)$ to $V_n(\cdot)$ in

$$V_n(\mathbf{b}) = \max_{a\in\mathcal{A}} \rho(\mathbf{b},a) + \min_{p_n^a\in\mathcal{P}^a} \sum_{s\in\mathcal{S}} \left(\mathbf{b}(s)\gamma \sum_{t,z\in\mathcal{S}\times\mathcal{Z}} p_n^a(t,z|s)V_{n+1}(\mathbf{b'}_{\mathbf{b},a,z})\right). \tag{101}$$

Let V and U be functions that map a belief state to a real number. For a fixed belief state, $\mathbf{b}$, let

$$a^* \equiv \operatorname*{argmax}_{a\in\mathcal{A}} \min_{p^a\in\mathcal{P}^a} \rho(\mathbf{b},a) + \sum_{s\in\mathcal{S}} \left(\mathbf{b}(s)\gamma \sum_{t,z\in\mathcal{S}\times\mathcal{Z}} p^a(t,z|s)V(\mathbf{b}'_{\mathbf{b},a,z})\right) \tag{102}$$

$$p^{a,*} \equiv \operatorname*{argmin}_{p^a\in\mathcal{P}^a} \rho(\mathbf{b},a) + \sum_{s\in\mathcal{S}} \left(\mathbf{b}(s)\gamma \sum_{t,z\in\mathcal{S}\times\mathcal{Z}} p^a(t,z|s)U(\mathbf{b}'_{\mathbf{b},a,z})\right), \forall a\in\mathcal{A}, \tag{103}$$

where $\mathbf{b}'_{\mathbf{b},a,z}$ is defined in (87). Suppose that $\mathcal{L}U(\mathbf{b}) \le \mathcal{L}V(\mathbf{b})$. Then,

$$0 \le \mathcal{L}V(\mathbf{b}) - \mathcal{L}U(\mathbf{b}) \tag{104}$$

$$\le \rho(\mathbf{b},a^*) + \sum_{s\in\mathcal{S}} \left(\mathbf{b}(s)\gamma \sum_{t,z\in\mathcal{S}\times\mathcal{Z}} p^{a^*,*}(t,z|s)V(\mathbf{b}'_{\mathbf{b},a^*,z})\right) \tag{105}$$

$$- \rho(\mathbf{b},a^*) + \sum_{s\in\mathcal{S}} \left(\mathbf{b}(s)\gamma \sum_{t,z\in\mathcal{S}\times\mathcal{Z}} p^{a^*,*}(t,z|s)U(\mathbf{b}'_{\mathbf{b},a^*,z})\right). \tag{106}$$

The second inequality remains valid because the belief-reward terms in (106) cancel. This leads to a simplification which does not include a reward term and the remainder of the proof follows directly from Osogami (2015, Theorem 2). □

2.2 Detailed Problem Formulation

2.2.1 Tiger Problem

The tiger problem is a classic POMDP introduced in Cassandra et al. (1994). This section describes the original nominal formulation and our robust formulation for an instance of size $|\mathcal{S}| = 2, |\mathcal{A}| = 3$, and $|\mathcal{Z}| = 2$.

Figures 31 and 32 show the hidden Markov models for the tiger problem dynamics, where solid lines indicate observations and dashed lines indicate hidden states for all hidden Markov model figures.

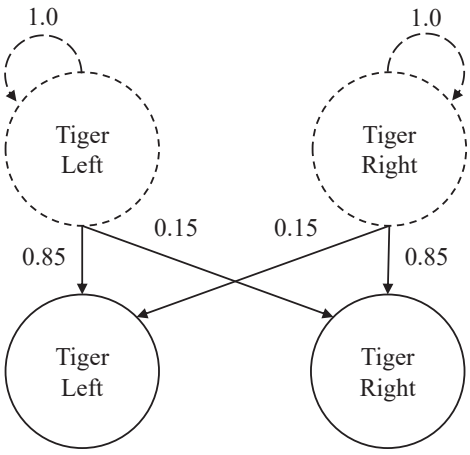

Figure 31. Tiger hidden Markov model (a = listen)

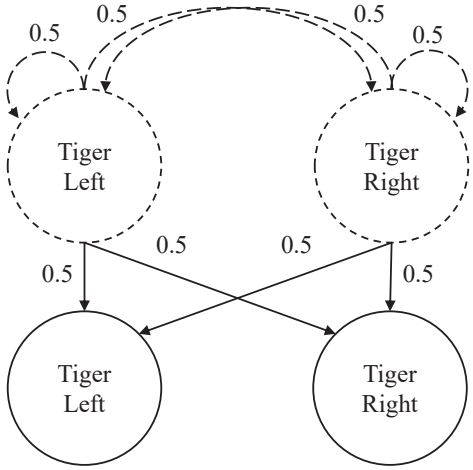

Figure 32. Tiger hidden Markov model, (a = open left or open right)

The formal POMDP components are defined as follows,

$$\mathcal{S} = \{\text{tiger-left}, \text{tiger-right}\} \tag{107}$$

$$\mathcal{A} = \{\text{listen}, \text{open-left}, \text{open-right}\} \tag{108}$$

$$\mathcal{Z} = \{\text{tiger-left}, \text{tiger-right}\} \tag{109}$$

$$T(s,a,s') = \begin{cases} 1, & s = s', a = \text{listen} \\ 0.5, & a \neq \text{listen}, s' = \text{tiger-left} \\ 0.5, & a \neq \text{listen}, s' = \text{tiger-right} \\ 0, & \text{otherwise} \end{cases} \tag{110}$$

$$O(a,s',z) = \begin{cases} 0.85, & s' = \text{tiger-left}, a = \text{listen}, z = \text{tiger-left} \\ 0.15, & s' = \text{tiger-left}, a = \text{listen}, z = \text{tiger-right} \\ 0.85, & s' = \text{tiger-right}, a = \text{listen}, z = \text{tiger-right} \\ 0.15, & s' = \text{tiger-right}, a = \text{listen}, z = \text{tiger-left} \\ 0.5, & a \neq \text{listen} \end{cases} \tag{111}$$

$$R(s,a) = \begin{cases} -100, & s = \text{tiger-left}, a = \text{open-left} \\ -100, & s = \text{tiger-right}, a = \text{open-right} \\ 10, & s = \text{tiger-left}, a = \text{open-right} \\ 10, & s = \text{tiger-right}, a = \text{open-left} \\ -1, & a = \text{listen} \end{cases} \tag{112}$$

$$\gamma = 0.95. \tag{113}$$

In the robust POMDP version of the problem, the transition function remains precise, but the observation function is defined using ambiguity sets rather than probability distributions. The ambiguity sets are defined as intervals around the nominal observation function using the ambiguity set half-width, δ, where $\delta \in \{0.001, 0.01, 0.1, 0.2\}$. The bounds of the ambiguity sets are truncated at ϵ and $1 - \epsilon$ for the following

observation function,

$$O^R(a, s', z) = \begin{cases} [0.85 - \delta, 0.85 + \delta], & s' = \text{tiger-left}, a = \text{listen}, z = \text{tiger-left} \\ [0.15 - \delta, 0.15 + \delta], & s' = \text{tiger-left}, a = \text{listen}, z = \text{tiger-right} \\ [0.85 - \delta, 0.85 + \delta], & s' = \text{tiger-right}, a = \text{listen}, z = \text{tiger-right} \\ [0.15 - \delta, 0.15 + \delta], & s' = \text{tiger-right}, a = \text{listen}, z = \text{tiger-left} \\ [0.5 - \delta, 0.5 + \delta], & a \neq \text{listen}. \end{cases} \tag{114}$$

2.2.2 Crying Baby Problem

The crying baby problem is an illustrative POMDP introduced in Kochenderfer (2015). This section describes the original nominal formulation and our robust formulation for an instance of size $|\mathcal{S}| = 2, |\mathcal{A}| = 2$, and $|\mathcal{Z}| = 2$. Figures 33 and 34 show the hidden Markov models for the two-observation baby problem dynamics.

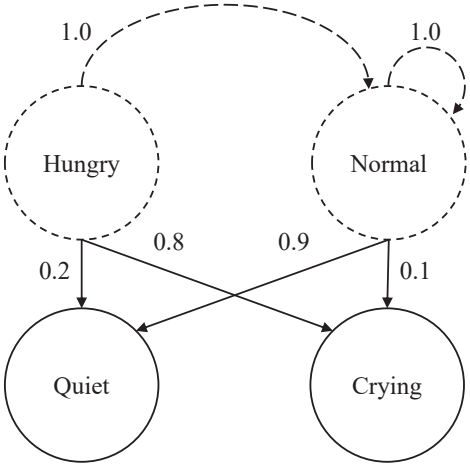

Figure 33. Baby hidden Markov model (a = feed)

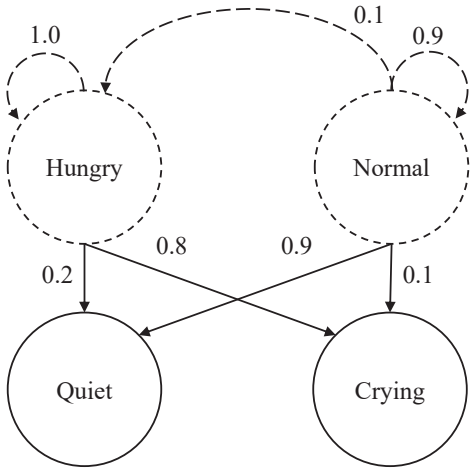

Figure 34. Baby hidden Markov model, (a = nothing)

The formal POMDP components are defined as follows,

$$\mathcal{S} = \{\text{hungry}, \text{normal}\} \tag{115}$$

$$\mathcal{A} = \{\text{feed}, \text{nothing}\} \tag{116}$$

$$\mathcal{Z} = \{\text{quiet}, \text{crying}\} \tag{117}$$

$$T(s, a, s') = \begin{cases} 1 & s = \text{hungry}, a = \text{nothing}, s' = \text{hungry} \\ 1 & a = \text{feed}, s' = \text{normal} \\ 0.1 & s = \text{normal}, a = \text{nothing}, s' = \text{hungry} \\ 0.9 & s = \text{normal}, a = \text{nothing}, s' = \text{normal} \\ 0 & \text{otherwise} \end{cases} \tag{118}$$

$$O(a, s', z) = \begin{cases} 0.8 & s' = \text{hungry}, z = \text{crying} \\ 0.2 & s' = \text{hungry}, z = \text{quiet} \\ 0.1 & s' = \text{normal}, z = \text{crying} \\ 0.9 & s' = \text{normal}, z = \text{quiet} \\ 0 & \text{otherwise} \end{cases} \tag{119}$$

$$R(s, a) = \begin{cases} -15 & s = \text{hungry}, a = \text{feed} \\ -10 & a = \text{hungry}, a = \text{nothing} \\ -5 & s = \text{normal}, a = \text{feed} \\ 0 & \text{otherwise} \end{cases} \tag{120}$$

$$\gamma = 0.9. \tag{121}$$

In the robust POMDP version of the problem, the transition function and the observation function are defined using ambiguity sets rather than probability distributions.

The ambiguity sets are defined as intervals around the nominal observation function using the ambiguity set half-width, δ, where $\delta \in \{0.001, 0.01, 0.1, 0.2\}$. The bounds of the ambiguity sets are truncated at $0 + \epsilon$ and $1 - \epsilon$ for the following transition and observation functions,

$$T^R(s, a, s') = \begin{cases} [1-\delta, 1+\delta] & s = \text{hungry}, a = \text{nothing}, s' = \text{hungry} \\ [1-\delta, 1+\delta] & a = \text{feed}, s' = \text{normal} \\ [0.1-\delta, 0.1+\delta] & s = \text{normal}, a = \text{nothing}, s' = \text{hungry} \\ [0.9-\delta, 0.9+\delta] & s = \text{normal}, a = \text{nothing}, s' = \text{normal} \\ [0-\delta, 0+\delta] & \text{otherwise} \end{cases} \tag{122}$$

$$O^R(s, a, z) = \begin{cases} [0.8-\delta, 0.8+\delta] & s = \text{hungry}, z = \text{crying} \\ [0.2-\delta, 0.2+\delta] & s = \text{hungry}, z = \text{quiet} \\ [0.1-\delta, 0.1+\delta] & s = \text{normal}, z = \text{crying} \\ [0.9-\delta, 0.9+\delta] & s = \text{normal}, z = \text{quiet} \\ [0-\delta, 0+\delta] & \text{otherwise.} \end{cases} \tag{123}$$

2.2.3 Rock Diagnosis Problem

The rock diagnosis problem is an information gathering belief-reward POMDP introduced in Araya-López (2013). This section describes the original nominal formulation and our robust formulation for an instance of size $|\mathcal{S}| = 4, |\mathcal{A}| = 3$, and $|\mathcal{Z}| = 3$. Figures 35, 36, and 37 show the hidden Markov models for the 2×1 rock diagnosis problem.

The formal POMDP components are defined as follows. Each state, $s \in \mathcal{S}$, includes a position component and a rock-type component. In this small problem, the 2×1 grid-world consists of two positions, 1 and 2, and two rock types, *bad* and *good*.

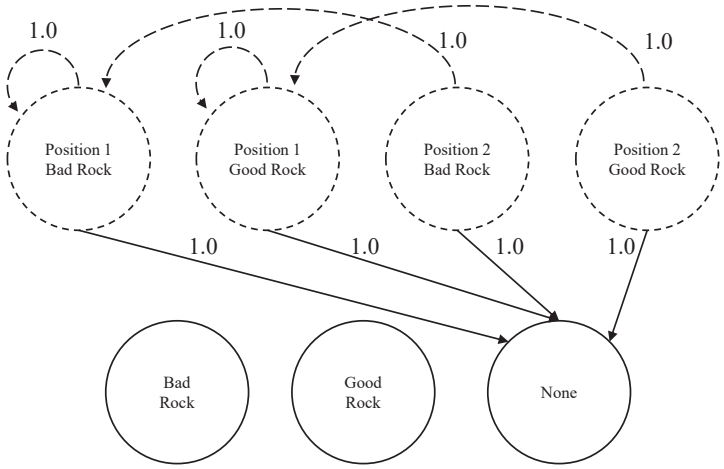

Figure 35. Rock diagnosis hidden Markov model (a = left)

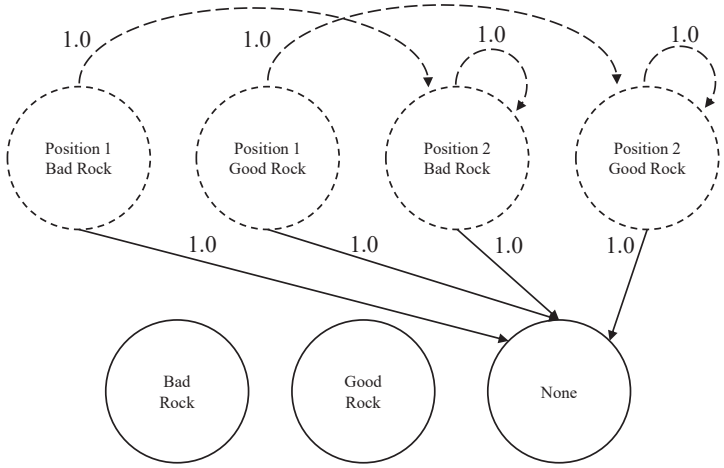

Figure 36. Rock diagnosis hidden Markov model (a = right)

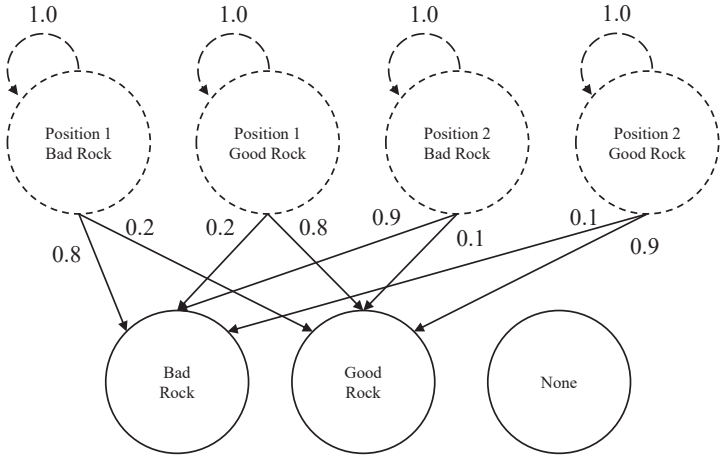

Figure 37. Rock diagnosis hidden Markov model, (a = check)

The rock is in position 2 and the sensor starts in position 1. Note that the position refers to the position of the sensor (the position of the rock is fixed and known to the decision-maker). The observation function depends on the parameters that control the rate at which the sensor accuracy decays with distance. In this case, $p_1 = 0.81$ and $p_2 = 0.90$ for the probability of correct detection when the sensor is in position 1 and position 2, respectively. The belief reward function is defined as a set of alpha vectors, Γ, where

$$\Gamma = \{[1, -1/3, -1/3, -1/3], \tag{124}$$

$$[-1/3, 1, -1/3, -1/3], \tag{125}$$

$$[-1/3, -1/3, 1, -1/3], \tag{126}$$

$$[-1/3, -1/3, -1/3, 1]\}. \tag{127}$$

This gives the following formulation for the nominal ρPOMDP,

$$\mathcal{S} = \{\text{1-bad}, \text{1-good}, \text{2-bad}, \text{2-good}\} \tag{128}$$

$$\mathcal{A} = \{\text{left}, \text{right}, \text{check}\} \tag{129}$$

$$\mathcal{Z} = \{\text{good}, \text{bad}, \text{none}\} \tag{130}$$

$$T(s, a, s') = \begin{cases} 1 & s = \text{1-bad}, a = \text{left}, s' = \text{1-bad} \\ 1 & s = \text{1-bad}, a = \text{right}, s' = \text{2-bad} \\ 1 & s = \text{1-bad}, a = \text{check}, s' = \text{1-bad} \\ 1 & s = \text{1-good}, a = \text{left}, s' = \text{1-good} \\ 1 & s = \text{1-good}, a = \text{right}, s' = \text{2-good} \\ 1 & s = \text{1-good}, a = \text{check}, s' = \text{1-good} \\ 1 & s = \text{2-bad}, a = \text{left}, s' = \text{1-bad} \\ 1 & s = \text{2-bad}, a = \text{right}, s' = \text{2-bad} \\ 1 & s = \text{2-bad}, a = \text{check}, s' = \text{2-bad} \\ 1 & s = \text{2-good}, a = \text{left}, s' = \text{1-good} \\ 1 & s = \text{2-good}, a = \text{right}, s' = \text{2-good} \\ 1 & s = \text{2-good}, a = \text{check}, s' = \text{2-good} \\ 0 & \text{otherwise} \end{cases} \tag{131}$$

$$O(a, s', z) = \begin{cases} 1 & a = \text{left}, z = \text{none} \\ 1 & a = \text{right}, z = \text{none} \\ p_1 & a = \text{check}, s' = \text{1-bad}, z = \text{bad} \\ 1 - p_1 & a = \text{check}, s' = \text{1-bad}, z = \text{good} \\ 1 - p_1 & a = \text{check}, s' = \text{1-good}, z = \text{bad} \\ p_1 & a = \text{check}, s' = \text{1-good}, z = \text{good} \\ p_2 & a = \text{check}, s' = \text{2-bad}, z = \text{bad} \\ 1 - p_2 & a = \text{check}, s' = \text{2-bad}, z = \text{good} \\ 1 - p_2 & a = \text{check}, s' = \text{2-good}, z = \text{bad} \\ p_2 & a = \text{check}, s' = \text{2-good}, z = \text{good} \\ 0 & \text{otherwise} \end{cases} \tag{132}$$

$$\rho(\mathbf{b}) = \max_{\alpha \in \Gamma} (\mathbf{b}\alpha) \tag{133}$$

$$\gamma = 0.9. \tag{134}$$

In the robust ρPOMDP version of the problem, the transition function remains precise, but the observation function is defined using ambiguity sets rather than probability distributions. The ambiguity sets are defined as intervals around the nominal observation function using two ambiguity set half-width parameters, $\delta_1 =$ 1e$-$6 and $\delta_2 = 0.4$ as follows. The variable-width ambiguity set models a trade-off between uncertainty and ambiguity with respect to sensor performance. The bounds

of the ambiguity sets are truncated at ϵ and $1-\epsilon$ for the following observation function,

$$O^R(a, s', z) = \begin{cases} 1 & a = \text{left}, z = \text{none} \\ 1 & a = \text{right}, z = \text{none} \\ [p_1 - \delta_1, p_1 + \delta_1] & a = \text{check}, s' = \text{1-bad}, z = \text{bad} \\ [1 - (p_1 + \delta_1), 1 - (p_1 - \delta_1)] & a = \text{check}, s' = \text{1-bad}, z = \text{good} \\ [1 - (p_1 + \delta_1), 1 - (p_1 - \delta_1)] & a = \text{check}, s' = \text{1-good}, z = \text{bad} \\ [p_1 - \delta_1, p_1 + \delta_1] & a = \text{check}, s' = \text{1-good}, z = \text{good} \\ [p_2 - \delta_2, p_2 + \delta_2] & a = \text{check}, s' = \text{2-bad}, z = \text{bad} \\ [1 - (p_2 + \delta_2), 1 - (p_2 - \delta_2)] & a = \text{check}, s' = \text{2-bad}, z = \text{good} \\ [1 - (p_2 + \delta_2), 1 - (p_2 - \delta_2)] & a = \text{check}, s' = \text{2-good}, z = \text{bad} \\ [p_2 - \delta_2, p_2 + \delta_2] & a = \text{check}, s' = \text{2-good}, z = \text{good} \\ 0 & \text{otherwise.} \end{cases} \tag{135}$$

Bibliography

ABC News (2007). Afghanistan - Where Things Stand. URL `https://abcnews.go.com/images/PollingUnit/1049a1Afghanistan-WhereThingsStand.pdf`.

Aberdeen, D. (2003). A (revised) survey of approximate methods for solving partially observable Markov decision processes. Technical report, National ICT Australia, Canberra.

Aghassi, M. and Bertsimas, D. (2006). Robust game theory. *Mathematical Programming*, 107(1):231–273.

Agra, A., Santos, M. C., Nace, D., and Poss, M. (2016). A dynamic programming approach for a class of robust optimization problems. *SIAM Journal on Optimization*, 26(3):1799–1823.

Ahipaşaoğlu, S. D., Meskarian, R., Magnanti, T. L., and Natarajan, K. (2015). Beyond normality: A cross moment-stochastic user equilibrium model. *Transportation Research Part B: Methodological*, 81:333–354. doi:10.1016/j.trb.2015.01.005.

Ahmed, A., Varakantham, P., Lowalekar, M., Adulyasak, Y., and Jaillet, P. (2017). Sampling based approaches for minimizing regret in uncertain Markov decision processes (MDPs). *Journal of Artificial Intelligence Research*, 59:229–264.

Ahner, D. K. and Parson, C. R. (2014). Optimal multi-stage allocation of weapons to targets using adaptive dynamic programming. *Optimization Letters*, 9(8):1689–1701. doi:10.1007/s11590-014-0823-x.

Alismail, F., Xiong, P., and Singh, C. (2018). Optimal wind farm allocation in multi-area power systems using distributionally robust optimization approach. *IEEE Transactions on Power Systems*, 33(1):536.

Alpern, S., Morton, A., and Papadaki, K. (2011). Patrolling games. *Operations Research*, 59(5):1246–1257. doi:10.1287/opre.1110.0983.

Aoki, M. (1965). Optimal control of partially observable Markovian systems. *Journal of The Franklin Institute*, 280(5):367–386.

Arad, A. and Gayer, G. (2012). Imprecise data sets as a source of ambiguity: A model and experimental evidence. *Management Science*, 58(1):188–202. doi:10.1287/mnsc.1110.1463.

Araya-López, M., Buffet, O., and Thomas, V. (2010). A POMDP extension with belief-dependent rewards. In *Advances in Neural Information Processing Systems*.

Araya-López, M. (2013). *Near-Optimal Algorithms for Sequential Information-Gathering Decision Problems*. PhD thesis, Universite de Lorraine.

Arnhart, L. M. B. and King, M. L. (2018). Are we there yet? Implementing best practices in assessments. *Military Review*, 97(3):20–29.

Arrow, K. J. and Hurwicz, L. (1972). An optimality criterion for decision making under ignorance. In Carter, C. and Ford, J., editors, *Uncertainty and Expectations in Economics*, pages 1–11 Basil Blackwell, Oxford.

Arslan, A., Kaminski, B., Prybol, C., Quinn, J., and Finnegan, R. (2018). DataFrames. URL `https://github.com/JuliaData/DataFrames.jl`.

Artzner, P., Delbaen, F., Eber, J. M., Heath, D., and Ku, H. (2007). Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Operations Research*, 152(1):5–22. doi:10.1007/s10479-006-0132-6.

Asanjarani, A., Nazarathy, Y., and Pollett, P. K. (2017). Parameter and state estimation in queues and related stochastic models: A bibliography. URL `https://people.smp.uq.edu.au/PhilipPollett/papers/Qest/QEstAnnBib.pdf`.

Astrom, K. J. (1965). Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10(1):174–205. doi:10.1016/0022-247X(65)90154-X.

Augustin, T., Coolen, F., de Cooman, G., and Troffaes, M., editors (2014). *Introduction to imprecise probability*. Wiley, Hoboken.

Bai, M. and Yang, Z. (2014). Distributionally robust self-scheduling optimization with CO2 emissions constraints under uncertainty of prices. *Journal of Applied Mathematics*.

Baillon, A., Huang, Z., Selim, A., and Wakker, P. P. (2018). Measuring ambiguity attitudes for all (natural) events. *Econometrica*, 86(5):1839–1858. doi:10.3982/ecta14370.

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, 27(1):173–196. doi:10.1257/jep.27.1.173.

Bazovkin, P. and Mosler, K. (2015). A general solution for robust linear programs with distortion risk constraints. *Annals of Operations Research*, 229(1):103–120. doi:10.1007/s10479-015-1786-8.

Beale, E. M. L. (1955). On minizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society Series B*, 17:173–184.

Bellman, R. E. and Zadeh, L. A. (1970). Decision-making in a fuzzy environment. *Management Science*, 17(4):B141–B164.

Ben-Haim, Y. (2006). *Info-Gap decision theory: Decisions under severe uncertainty*. Academic Press, Oxford.

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust optimization*. Princeton University Press, Princeton.

Ben-Tal, A., Bertsimas, D., and Brown, D. B. (2010). A soft robust model for optimization under ambiguity. *Operations Research*, 58(4-part-2):1220–1234. doi:10.1287/opre.1100.0821.

Ben-Tal, A., den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357. doi:10.1287/mnsc.1120.1641.

Berger, J. O., Moreno, E., Pericchi, L. R., Bayarri, M. J., Bernardo, J. M., Cano, J. A., la Horra, J., Martin, J., Rios-Insúa, D., and Betrò, B. (1994). An overview of robust Bayesian analysis. *Test*, 3(1):5–124.

Bertsimas, D. and Brown, D. B. (2009). Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6):1483–1495. doi:10.1287/opre.1080.0646.

Bertsimas, D. and Goyal, V. (2010). On the power of robust solutions in two-stage stochastic and adaptive optimization problems. *Mathematics of Operations Research*, 35(2):284–305. doi:10.1287/moor.1090.0440.

Bertsimas, D. and Servi, L. D. (1992). Deducing queueing from transactional data: The queue inference engine, revisited. *Operations Research*, 40(3-supplement-2): 217–228.

Bertsimas, D. and Sim, M. (2004). The price of robustness. *Operations Research*, 52 (1):35–53. doi:10.1287/opre.1030.0065.

Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501.

Bertsimas, D., Gupta, V., and Kallus, N. (2018a). Data-driven robust optimization. *Mathematical Programming*, 167(2):235–292. doi:10.1007/s10107-017-1125-8.

Bertsimas, D., Sim, M., and Zhang, M. (2018b). Adaptive distributionally robust optimization. *Management Science*, 65(2):604–618.

Bertuccelli, L. F., Bethke, B., and How, J. P. (2009). Robust adaptive Markov decision processes in multi-vehicle applications. In *2009 American Control Conference*, pages 1304–1309. doi:10.1109/ACC.2009.5160511.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98. doi:10.1137/141000671.

Binmore, K., Stewart, L., and Voorhoeve, A. (2012). How much ambiguity aversion? Finding indifferences between Ellsberg's risky and ambiguous bets. *Journal of Risk and Uncertainty*, 45(3):215–238.

Boardman, N. T., Lunday, B. J., and Robbins, M. J. (2017). Heterogeneous surface-to-air missile defense battery location: A game theoretic approach. *Journal of Heuristics*, 23(6):417–447.

Boloori, A. and Cook, C. B. (2017). Data-Driven management of post-transplant medications: An APOMDP approach. *HKS Working Paper RWP17-036*.

Bosansky, B., Kiekintveld, C., Lisy, V., and Pěchoucek, M. (2014). An exact double-oracle algorithm for zero-sum extensive-form games with imperfect information. *Journal of Artificial Intelligence Research*, 51:829–866. doi:10.1002/bkcs.10128.

Brown, N. and Sandholm, T. (2018). Solving imperfect-information games via discounted regret minimization. URL `http://arxiv.org/abs/1809.04040`.

Brown, N., Lerer, A., Gross, S., and Sandholm, T. (2018). Deep counterfactual regret minimization. URL `https://arxiv.org/abs/1710.11424`.

Burnetas, A. and Economou, A. (2007). Equilibrium customer strategies in a single server Markovian queue with setup times. *Queueing Systems*, 56(3-4):213–228. doi:10.1007/s11134-007-9036-7.

Camerer, C. and Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, 5(4):325–370.

Cassandra, A., Kaelbling, L., and Littman, M. (1994). Acting optimaly in partially observable stochastic domains. In *Proceedings 12th National Conference on Artificial Intelligence (AAAI-94)*, volume 94, pages 1023–1028.

Cassandra, A., Littman, M., and Zhang, N. (1997). Incremental pruning: A simple, fast, exact method for partially observable Markov decision processes. In *Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, pages 54–61.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., and Montrucchio, L. (2011). Uncertainty averse preferences. *Journal of Economic Theory*, 146(4):1275–1330. doi:10.1016/j.jet.2011.05.006.

Chateauneuf, A., Eichberger, J., and Grant, S. (2007). Choice under uncertainty with the best and worst in mind: Neo-additive capacities. *Journal of Economic Theory*, 137(1):538–567.

Chen, L. and Leneutre, J. (2009). A game theoretical framework on intrusion detection in heterogeneous networks. *IEEE Transactions on Information Forensics and Security*, 4(2):165–178. doi:10.1109/TIFS.2009.2019154.

Chen, Z., Yu, P., and Haskell, W. B. (2018). Distributionally robust optimization for sequential decision making. *arXiv Preprint*.

Choquet, G. (1954). Theory of capacities. In *Annales de l'Institut Fourier*, volume 5, pages 131–295.

Claßen, G., Koster, A. M. C., and Schmeink, A. (2015). The multi-band robust knapsack problem - a dynamic programming approach. *Discrete Optimization*, 18: 123–149.

Conflict Casualties Monitor (2018). Iraq Body Count. URL `https://www.iraqbodycount.org/database/`.

Conlisk, J. (1996). Why bounded rationality? *Journal of Economic Literature*, 34 (2):669–700.

Cook, T. D., Campbell, D. T., and Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin, Boston.

Cousins, J. B. and Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14(4):397–418. doi:10.3102/01623737014004397.

Cox, L. A. T. (2012). Confronting deep uncertainties in risk analysis. *Risk Analysis*, 32(10):1607–1629. doi:10.1111/j.1539-6924.2012.01792.x.

Dantzig, G. B. (1955). Linear programming under uncertainty. *Management Science*, 1(3-4):197–206.

Davis, M. T., Robbins, M. J., and Lunday, B. J. (2017). Approximate dynamic programming for missile defense interceptor fire control. *European Journal of Operational Research*, 259(3):873–886. doi:10.1016/j.ejor.2016.11.023.

Davis, T., Waugh, K., and Bowling, M. (2018). Solving large extensive-form games with strategy constraints. URL `https://arxiv.org/abs/1809.07893`.

de Finetti, B. (1974). *Theory of probability: A critical introductory treatment*. John Wiley & Sons, New York.

Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3): 595–612. doi:10.1287/opre.1090.0741.

Delgado, K. V., de Barros, L. N., Cozman, F. G., and Sanner, S. (2011a). Using mathematical programming to solve factored Markov decision processes with imprecise probabilities. *International Journal of Approximate Reasoning*, 52(7):1000–1017. doi:10.1016/j.ijar.2011.04.002.

Delgado, K. V., Sanner, S., and de Barros, L. N. (2011b). Efficient solutions to factored MDPs with imprecise transition probabilities. *Artificial Intelligence*, 175 (9-10):1498–1527. doi:10.1016/j.artint.2011.01.001.

Delgado, K. V., de Barros, L. N., Dias, D. B., and Sanner, S. (2016). Real-time dynamic programming for Markov decision processes with imprecise probabilities. *Artificial Intelligence*, 230:192–223. doi:10.1016/j.artint.2015.09.005.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 458(2):325–339. doi:10.1016/j.jmaa.2017.10.006.

Department of the Army (2014). Field Manual 6-0: Commander and Staff Organization and Operations.

Department of the Army (2017). Army Doctrine Publication 3-0 Operations.

Destercke, S., Dubois, D., and Chojnacki, E. (2008). Unifying practical uncertainty representations - I: Generalized p-boxes. *International Journal of Approximate Reasoning*, 49(3):649–663. doi:10.1016/j.ijar.2008.07.003.

Diecidue, E. and Wakker, P. (2001). On the intuition of rank-dependent utility. *Journal of Risk and Uncertainty*, 23(3):281–298.

Dimitrov, N. B., Dimitrov, S., and Chukova, S. (2014). Robust decomposable Markov decision processes motivated by allocating school budgets. *European Journal of Operational Research*, 239(1):199–213. doi:10.1016/j.ejor.2014.05.003.

Dimitrova, R., Fu, J., and Topcu, U. (2016). Robust optimal policies for Markov decision processes with safety-threshold constraints. In *IEEE 55th Conference on Decision and Control*, pages 7081–7086. doi:10.1109/CDC.2016.7799360.

Dimmock, S. G., Kouwenberg, R., Mitchell, O. S., and Peijnenburg, K. (2015a). Estimating ambiguity preferences and perceptions in multiple prior models: Evidence from the field. *Journal of Risk and Uncertainty*, 51(3):219–244. doi:10.1007/s11166-015-9227-2.

Dimmock, S. G., Kouwenberg, R., and Wakker, P. P. (2015b). Ambiguity attitudes in a large representative sample. *Management Science*, 62(5):1363–1380. doi:10.2139/ssrn.1876580.

Dong, J. and Whitt, W. (2015). Stochastic grey-box modeling of queueing systems: fitting birth-and-death processes to data. *Queueing Systems*, 79(3-4):391–426.

Doria, S. (2017). On the disintegration property of coherent upper conditional prevision defined by the Choquet integral with respect to its associated Hausdorff outer measure. *Annals of Operations Research*, 256(2):253–269. doi:10.1007/s10479-016-2270-9.

Dubois, D. and Prade, H. (2003). Possibility theory and its applications: A retrospective and prospective view. In *The 12th IEEE International Conference on Fuzzy Systems*, pages 5–11. doi:10.1109/FUZZ.2003.1209314.

Dubois, D. and Prade, H. (1987). Properties of measures of information in evidence and possibility theories. *Fuzzy Sets and Systems*, 24(2):161–182.

Dubois, D. and Prade, H. (2001). Possibility theory, probability theory and multiple-valued logics: A clarification. *Annals of Mathematics and Artificial Intelligence*, 32 (1-4):35–66. doi:10.1023/A:1016740830286.

Dubois, D., Prade, H., and Sabbadin, R. (2001). Decision-theoretic foundations of qualitative possibility theory. *European Journal of Operational Research*, 128(3): 459–478. doi:10.1016/S0377-2217(99)00473-7.

Dubra, J., Maccheroni, F., and Ok, E. A. (2004). Expected utility theory without the completeness axiom. *Journal of Economic Theory*, 115(1):118–133. doi:10.1016/S0022-0531(03)00166-2.

Dupacova, J., Growe-Kuska, N., and Romisch, W. (2003). Scenario reduction in stochastic programming. *Mathematical Programming, Series A*, 95:493–511. doi:10.1007/s10107-002-0331-0.

Dynkin, E. (1965). *Markov processes*. Springer, Berlin. doi:10.1007/978-3-662-00031-1.

Eck, A. and Soh, L. K. (2012). Evaluating POMDP rewards for active perception. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 3*, pages 1221–1222.

Edelson, N. M. and Hilderbrand, D. K. (1975). Congestion tolls for Poisson queueing process. *Econometrica*, 43(1):81–92.

Egorov, M., Sunberg, Z. N., Balaban, E., Wheeler, T. A., Gupta, J. K., and Kochenderfer, M. J. (2017). POMDPs.jl: A framework for sequential decision making under uncertainty. *Journal of Machine Learning Research*, 18(26):1–5.

Eichberger, J. and Kelsey, D. (1999). E-capacities and the Ellsberg paradox. *Theory and Decision*, 46(2):107–140. doi:10.1023/A:1004994630014.

Eisner, E. W. (1976). Educational connoisseurship and criticism: Their form and functions in educational evaluation. *Journal of Aesthetic Education*, 10(3/4):135.

Eliaz, K. and Ok, E. A. (2006). Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences. *Games and Economic Behavior*, 56(1):61–86. doi:10.1016/j.geb.2005.06.007.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669.

Epstein, L. and Schneider, M. (2007). Learning under ambiguity. *The Review of Economic Studies*, 74(4):1275–1303.

Ergin, H. and Gul, F. (2009). A theory of subjective compound lotteries. *Journal of Economic Theory*, 144(3):899–929. doi:10.1016/j.jet.2008.08.003.

Etner, J., Jeleva, M., and Tallon, J. M. (2012). Decision theory under ambiguity. *Journal of Economic Surveys*, 26(2):234–270. doi:10.1111/j.1467-6419.2010.00641.x.

Farina, G., Kroer, C., and Sandholm, T. (2017). Regret minimization in behaviorally-constrained zero-sum games. URL `http://arxiv.org/abs/1711.03441`.

Fehr, E. and Rangel, A. (2011). Neuroeconomic foundations of economic choice - recent advances. *Journal of Economic Perspectives*, 25(4):3–30. doi:10.1257/jep.25.4.3.

Ferson, S., Joslyn, C. A., Helton, J. C., Oberkampf, W. L., and Sentz, K. (2004). Summary from the epistemic uncertainty workshop: Consensus amid diversity. *Reliability Engineering and System Safety*, 85(1-3):355–369. doi:10.1016/j.ress.2004.03.023.

Fine, T. L. (1977). Review: Glenn Shafer, A mathematical theory of evidence. *Bulletin of the American Mathematical Society*, 83(4):667–672. doi:10.1093/ije/6.1.83.

Fishburn, P. C. (1986). The axioms of subjective probability. *Statistical Science*, 1 (3):335–345. doi:10.1214/ss/1177013611.

Fitzpatrick, J. L., Sanders, J. R., and Worthen, B. R. (2011). *Program Evaluation: Alternative Approaches and Practical Guidelines*. Pearson, Boston, 4th edition.

Frey, J. C. and Kaplan, E. H. (2010). Queue inference from periodic reporting data. *Operations Research Letters*, 38(5):420–426. doi:10.1016/j.orl.2010.05.005.

Fussuma, F. L., Delgado, K. V., and de Barros, L. N. (2014). B2RTDP: An efficient solution for Bounded-Parameter Markov Decision Process. In *2014 Brazilian Conference on Intelligent Systems*, pages 128–133.

Gabrel, V., Murat, C., and Thiele, A. (2014). Recent advances in robust optimization: An overview. *European Journal of Operational Research*, 235(3):471–483. doi:10.1016/j.ejor.2013.09.036.

Gagnon, M. N., Truelove, J., Kapadia, A., Haines, J., and Huang, O. (2010). Towards net-centric cyber survivability for ballistic missile defense. In *International Symposium on Architecting Critical Systems*, pages 125–141, Berlin Springer. doi:10.1007/978-3-642-13556-9_8.

Gajdos, T., Hayashi, T., Tallon, J. M., and Vergnaud, J. C. (2008). Attitude toward imprecise information. *Journal of Economic Theory*, 140(1):27–65. doi:10.1016/j.jet.2007.09.002.

Ganzfried, S. and Sun, Q. (2018). Bayesian opponent exploitation in imperfect-information games. In *IEEE Conference on Computatonal Intelligence and Games, CIG*, pages 1–18. doi:10.1109/CIG.2018.8490452.

Ganzfried, S. and Sandholm, T. (2015). Safe opponent exploitation. *ACM Transactions on Economics and Computation*, 3(2):8:1 – 8:28. doi:10.1016/j.apcatb.2006.06.014.

Ghirardato, P., Maccheroni, F., and Marinacci, M. (2004). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, 118(2):133–173.

Giang, P. H. and Shenoy, P. P. (2005). Two axiomatic approaches to decision making using possibility theory. *European Journal of Operational Research*, 162(2):450–467. doi:10.1016/j.ejor.2003.05.004.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153. doi:10.1016/0304-4068(89)90018-9.

Givan, R., Leach, S., and Dean, T. (2000). Bounded-Parameter Markov decision processes. *Artificial Intelligence*, 122(1-2):71–109.

Goerigk, M. and Schöbel, A. (2016). Algorithm engineering in robust optimization. In *Algorithm Engineering*, pages 245–279 Springer, Cham.

Good, I. J. (1977). Bruno de Finetti, Theory of Probability. *Bulletin of the American Mathematical Society*, 83(1):94–97.

Gotoh, J. Y. and Uryasev, S. (2017). Support vector machines based on convex risk functions and general norms. *Annals of Operations Research*, 249(1-2):301–328. doi:10.1007/s10479-016-2326-x.

Gotoh, J. Y., Kim, M. J., and Lim, A. E. B. (2017). Calibration of distributionally robust empirical optimization models. *arXiv Preprint*.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica*, 52(3):701–720.

Gross, D., Shortle, J. F., Thompson, J. M., and Harris, C. M. (2011). *Fundamentals of Queueing Theory*. John Wiley & Sons, Hoboken.

Guha, D., Goswami, V., and Banik, A. D. (2016). Algorithmic computation of steady-state probabilities in an almost observable GI/M/c queue with or without vacations under state dependent balking and reneging. *Applied Mathematical Modelling*, 40 (5-6):4199–4219. doi:10.1016/j.apm.2015.11.018.

Gul, F. and Pesendorfer, W. (2006). Random expected utility. *Econometrica*, 74(1): 121–146.

Guo, H., Goldsman, D., Tsui, K. L., Zhou, Y., and Wong, S. Y. (2016). Using simulation and optimisation to characterise durations of emergency department service times with incomplete data. *International Journal of Production Research*, 54(21):6494–6511. doi:10.1080/00207543.2016.1205760.

Guo, P. (2011). One-shot decision theory. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, 41(5):917–926. doi:10.1109/TSMCA.2010.2093891.

Guo, P. (2019). Focus theory of choice and its application to resolving the St. Petersburg, Allais, and Ellsberg paradoxes and other anomalies. *European Journal of Operational Research*, 276(3):1034–1043. doi:10.1016/j.ejor.2019.01.019.

Guo, P. and Li, Y. (2014). Approaches to multistage one-shot decision making. *European Journal of Operational Research*, 236(2):612–623. doi:10.1016/j.ejor.2013.12.038.

Guo, P. and Tanaka, H. (2010). Decision making with interval probabilities. *European Journal of Operational Research*, 203(2):444–454. doi:10.1016/j.ejor.2009.07.020.

Gutin, E., Kuhn, D., and Wiesemann, W. (2015). Interdiction games on Markovian PERT networks. *Management Science*, 61(5):999–1017. doi:10.1287/mnsc.2014.1973.

Hahn, E. M., Hashemi, V., Hermanns, H., Lahijanian, M., and Turrini, A. (2017). Multi-objective robust strategy synthesis for interval Markov decision processes. In Bertrand, N. and Bortolussi, L., editors, *Quantitative Evaluation of Systems*, pages 207–223 Springer, Cham. doi:10.1007/978-3-319-66335-7_13.

Hahn, G. J. and Meeker, W. Q. (2011). *Statistical intervals: A guide for practitioners*. John Wiley & Sons, Hoboken.

Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika*, 67(1):133–143. doi:10.1093/biomet/67.1.133.

Han, C. Y., Lunday, B. J., and Robbins, M. J. (2016). A game theoretic model for the optimal location of Integrated Air Defense System missile batteries. *INFORMS Journal on Computing*, 28(3):405–416. doi:10.1287/ijoc.2016.0690.

Hansen, E. A., Bernstein, D. S., and Zilberstein, S. (2004). Dynamic programming for partially observable stochastic games. In *Proceedings of the 19th National Conference on Artificial Intelligence (AAAI-04)*, pages 709–715.

Hansen, E. A. (1997). An improved policy iteration algorithm for partially observable MDPs. In *Advances in Neural Information Processing Systems*, pages 1015–1021.

Hansen, L. P. and Sargent, T. J. (2001). Robust control and model uncertainty. *American Economic Review*, 91(2):60–66.

Harless, D. W. and Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62(6):1251–1289.

Harmanec, D. (2002). Generalizing Markov decision processes to imprecise probabilities. *Journal of Statistical Planning and Inference*, 105(1):199–213.

Haskell, W. B., Kar, D., Fang, F., Tambe, M., Cheung, S., and Denicola, E. (2014). Robust protection of fisheries with COmPASS. In *Proceedings of the 26th Annual Conference on Innovative Applications of Artificial Intelligence*, pages 2978–2983.

Hazen, G. (1989). Ambiguity aversion and ambiguity content in decision making under uncertainty. *Annals of Operations Research*, 19(1):415–433.

Horak, K., Bosansky, B., and Pechoucek, M. (2017). Heuristic search value iteration for one-sided partially observable stochastic games. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI-17)*, pages 558–564.

Hu, Z. and Hong, L. (2012). Kullback-Leibler divergence constrained distributionally robust optimization. *Optimization Online Preprint*.

Huang, J., Zhou, K., and Guan, Y. (2017). A study of distributionally robust multistage stochastic optimization. *arXiv Preprint*.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.

Iancu, D. A. and Trichakis, N. (2014). Pareto efficiency in robust optimization. *Management Science*, 60(1):130–147. doi:10.1287/mnsc.2013.1753.

Itoh, H. and Nakamura, K. (2007). Partially observable Markov decision processes with imprecise parameters. *Artificial Intelligence*, 171(8-9):453–490.

Iyengar, G. N. (2005). Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280.

Jaffray, J. Y. (1989). Linear utility theory for belief functions. *Operations Research Letters*, 8(2):107–112.

Jain, M., Conitzer, V., and Tambe, M. (2013). Security scheduling for real-world networks. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems*, pages 215–222.

Jakubovskis, A. (2017). Strategic facility location, capacity acquisition, and technology choice decisions under demand uncertainty: Robust vs. non-robust optimization approaches. *European Journal of Operational Research*, 260(3):1095–1104.

Jech, T. (1992). The logarithmic distribution of leading digits and finitely additive measures. *Discrete Mathematics*, 108(1-3):53–57. doi:10.1016/0012-365X(92)90659-4.

Jensen, F. V. and Nielsen, T. D. (2013). Probabilistic decision graphs for optimization under uncertainty. *Annals of Operations Research*, 204(1):223–248. doi:10.1007/s10479-012-1263-6.

Jiang, H., Netessine, S., and Savin, S. (2011). Robust newsvendor competition under asymmetric information. *Operations research*, 59(1):254–261. doi:10.1287/opre.1100.0858.

Johanson, M. (2016). *Robust strategies and counter-strategies: From superhuman to optimal play*. PhD thesis, University of Alberta.

Johanson, M. and Bowling, M. (2009). Data biased robust counter strategies. *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS)*, page 264271.

Johanson, M., Zinkevich, M., and Bowling, M. (2008). Computing robust counter-strategies. In *Advances in Neural Information Processing Systems*, pages 721–728.

Johanson, M., Bard, N., Lanctot, M., Gibson, R., and Bowling, M. (2012). Efficient Nash equilibrium approximation through Monte Carlo counterfactual regret minimization. *AAMAS '12 Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, pages 837–846. doi:10.1073/pnas.202471199.

Joint Staff J-7 (2011). Commanders Handbook for Assessment Planning and Execution.

Jones, L. K. (1999). Inferring balking behavior from transactional data. *Operations Research*, 47(5):778–784. doi:10.1287/opre.47.5.778.

Jones, L. K. (2012). Remarks on queue inference from departure data alone and the importance of the queue inference engine. *Operations Research Letters*, 40(6): 503–505. doi:10.1016/j.orl.2012.08.001.

Jones, L. K. and Larson, R. C. (1995). Efficient computation of probabilities of events described by order statistics and applications to queue inference. *ORSA Journal on Computing*, 7(1):899–1499. doi:10.1287/ijoc.7.1.89.

JuliaStats (2018a). Distributions. URL `https://github.com/JuliaStats/Distributions.jl`.

JuliaStats (2018b). StatsBase. URL `https://github.com/JuliaStats/StatsBase.jl`.

Julien, B. (1994). An extension to possibilistic linear programming. *Fuzzy Sets and Systems*, 64(2):195–206.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263.

Kaplan, E. H. (2010). Terror queues. *Operations Research*, 58(4):773–784.

Kar, D., Nguyen, T. H., Fang, F., Brown, M., Sinha, A., Tambe, M., and Jiang, A. X. (2016). Trends and Applications in Stackelberg Security Games. In Basar, T. and Zaccour, G., editors, *Handbook of Dynamic Game Theory* Springer, Cham.

Kardeş, E. (2005). Robust stochastic games and applications to counter-terrorism strategies. Technical report, CREATE.

Kardeş, E. (2014). On discounted stochastic games with incomplete information on payoffs and a security application. *Operations Research Letters*, 42(1):7–11. doi:10.1016/j.orl.2013.10.005.

Kardeş, E., Ordóñez, F., and Hall, R. W. (2011). Discounted robust stochastic games and an application to queueing control. *Operations Research*, 59(2):365–382. doi:10.1287/opre.1110.0931.

Kataoka, S. (2016). A stochastic programming model. *Econometrica*, 31(1):181–196.

Kaufman, D. L. and Schaefer, A. J. (2013). Robust modified policy iteration. *INFORMS Journal on Computing*, 25(3):396–410.

Kennedy, D. P. (1972). A note on the number of busy servers in a GI/G/s queue in light traffic. *Journal of Applied Probability*, 9(4):868–869.

Kiefer, J. D. and Wolfowitz, J. (1955). On the theory of queues with many servers. *Transactions of the American Mathematical Society*, 78(1):1–18.

Kiennert, C., Ismail, Z., Debar, H., and Leneutre, J. (2018). A survey on game theoretic approaches for intrusion detection and response optimization. *ACM Computing Surveys*, 51(5):90:1 – 90:31.

Killian, T. W., Daulton, S., Konidaris, G., Doshi-Velez, F., and Konidaris, G. (2017). Robust and efficient transfer learning with hidden parameter Markov decision processes. In *Advances in Neural Information Processing Systems*, pages 6251–6262. doi:10.1038/nature14236.

Klibanoff, P., Marinacci, M., Applicata, M., and Torino, U. (2005). A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892.

Klibanoff, P., Marinacci, M., Alberto, C. C., and Torino, U. (2009). Recursive smooth ambiguity preferences. *Journal of Economic Theory*, 144(3):930–976.

Kline, A., Ahner, D., and Lunday, B. J. (2017). Real-time heuristic algorithms for the static weapon target assignment problem. *Journal of Heuristics*, pages 1–21.

Kline, A., Ahner, D., and Hill, R. (2018). The weapon-target assignment problem. *Computers and Operations Research*, 105:226–236. doi:10.1016/j.cor.2018.10.015.

Knight, F. H. (1921). *Risk, uncertainty and profit*. Hart, Schaffner and Marx, New York.

Kochenderfer, M. J. (2015). *Decision Making Under Uncertainty: Theory and Application*. MIT Press, Cambridge.

Koller, D., Megiddo, N., and von Stengel, B. (1994). Fast algorithms for finding randomized strategies in game trees. In *Proceedings of the 26th ACM Symposium on Theory of Computing*, pages 750–760.

Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin.

Kolmogorov, A. (2018). *Foundations of the theory of probability*. Dover, New York, 2nd edition.

Korzhyk, D., Yin, Z., Kiekintveld, C., Conitzer, V., and Tambe, M. (2011). Stackelberg vs. Nash in security games: An extended investigation of interchangeability, equivalence, and uniqueness. *Journal of Artificial Intelligence Research*, 41:297–327. doi:10.1613/jair.3269.

Kothiyal, A., Spinu, V., and Wakker, P. P. (2011). Prospect theory for continuous distributions: A preference foundation. *Journal of Risk and Uncertainty*, 42(3): 195–210. doi:10.1007/s11166-011-9118-0.

Krajbich, I., Oud, B., and Fehr, E. (2014). Benefits of neuroeconomic modeling: New policy interventions and predictors of preference. *American Economic Review*, 104 (5):501–506.

Kroer, C., Farina, G., and Sandholm, T. (2018). Solving large sequential games with the excessive gap technique. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*.

Kuhn, H. W. (1950). A simplified two-person poker. *Contributions to the Theory of Games*, 1:97–103.

Kumar, A. and Zilberstein, S. (2009). Dynamic programming approximations for partially observable stochastic games. In *22nd International FLAIRS Conference*, pages 547–552.

Kurniawati, H., Hsu, D., and Lee, W. S. (2008). SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In *Proceedings of Robotics: Science and Systems IV*.

Lambert, D. (1992). Zero-Inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14. doi:10.1080/00401706.1992.10485228.

Lang, M. (2017). First-order and second-order ambiguity aversion. *Management Science*, 63(4):1254–1269.

Larson, R. C. (1990). The queue inference engine: Deducing queue statistics from transactional data. *Management Science*, 36(5):586–601.

Lauwens, B. (2018). SimJulia. URL `https://github.com/BenLauwens/SimJulia.jl`.

Lei, C., Lin, W. H., and Miao, L. (2016). A two-stage robust optimization approach for the mobile facility fleet sizing and routing problem under uncertainty. *Computers & Operations Research*, 67:75–89.

Lempert, R. J. (2003). *Shaping the next one hundred years: New methods for quantitative, long-term policy analysis*. Rand Corporation, Santa Monica.

Lessin, A. M., Lunday, B. J., and Hill, R. R. (2018). A bilevel exposure-oriented sensor location problem for border security. *Computers and Operations Research*, 98:56–68. doi:10.1016/j.cor.2018.05.017.

Levin, J. (2006). *Choice Under Uncertainty*. Lecture Notes.

Li, B. and Si, J. (2010). Approximate robust policy iteration using multilayer perceptron neural networks for discounted infinite-horizon Markov decision processes with uncertain correlated transition matrices. *IEEE Transactions on Neural Networks*, 21(8):1270–1280.

Li, Z., Müller, J., Wakker, P. P., and Wang, T. V. (2017). The rich domain of ambiguity explored. *Management Science*, 64(7):3227–3240. doi:10.1287/mnsc.2017.2777.

Lim, A. E. B., Shanthikumar, J. G., and Shen, Z. J. (2006). Model uncertainty, robust optimization, and learning. *Tutorials in Operations Research: Models, Methods, and Applications for Innovative Decision Making*, pages 66–94. doi:10.1287/educ.1063.0023.

Lim, S. H., Xu, H., and Mannor, S. (2016). Reinforcement learning in robust Markov decision processes. *Mathematics of Operations Research*, 41(4):1325–1353.

Lisy, V., Davis, T., and Bowling, M. (2016). Counterfactual regret minimization in sequential security games. In *Proceedings of the 30th Conference on Artificial Intelligence (AAAI-16)*, pages 544–550.

Liu, J., Jin, X., Wang, T., and Yuan, Y. (2015). Robust multi-period portfolio model based on prospect theory and ALMV-PSO algorithm. *Expert Systems with Applications*, 42(20):7252–7262.

Liu, Y., Xu, H., Yang, S. J. S., and Zhang, J. (2018). Distributionally robust equilibrium for continuous games: Nash and Stackelberg models. *European Journal of Operational Research*, 265(2):631–643. doi:10.1016/j.ejor.2017.07.050.

Loizou, N. (2016). Distributionally robust games with risk-averse players. *arXiv Preprint*. doi:10.5220/0005753301860196.

Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368):805–824.

Loomes, G. and Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, 39(3-4):641–648. doi:10.1016/0014-2921(94)00071-7.

Lougee-Heimer, R. (2003). The Common Optimization INterface for Operations Research. *IBM Journal of Research and Development*, 47(1):57–66.

Lovejoy, W. S. (1991). A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, 28(1):47–66.

Lubin, M. and Dunning, I. (2015). Computing in operations research using Julia. *INFORMS Journal on Computing*, 27(2):238–248.

Lye, K. W. and Wing, J. M. (2005). Game strategies in network security. *International Journal of Information Security*, 4(1-2):71–86. doi:10.1007/s10207-004-0060-x.

Maccheroni, F., Marinacci, M., and Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6): 1447–1498. doi:10.1111/j.1468-0262.2006.00716.x.

Machina, M. J. (1982). "Expected utility" analysis without the independence axiom. *Econometrica*, 50(2):277–323.

Maiers, J. and Sherif, Y. S. (1985). Applications of fuzzy set theory. *IEEE Transactions on Systems, Man and Cybernetics*, 15(1):175–189. doi:10.1109/TSMC.1985.6313408.

Mannor, S., Mebel, O., and Xu, H. (2016). Robust MDPs with k-rectangular uncertainty. *Mathematics of Operations Research*, 41(4):1484–1509.

Marinacci, M. (2015). Model uncertainty. *Journal of the European Economic Association*, 13(6):1022–1100. doi:10.1111/jeea.12164.

Mark, M. M. and Shotland, R. L. (1985). Stakeholder-based evaluation and value judgments: The role of perceived power and legitimacy in the selection of stakeholder groups. *Evaluation Review*, 9(5):605–626.

Marquis, J. P., Mcnerney, M. J., Zimmerman, S. R., Archer, M., Boback, J., and Stebbins, D. (2016). Developing an assessment, monitoring, and evaluation framework for U.S. Department of Defense security cooperation. Technical report, RAND, Santa Monica.

Mathison, S. (2005). *Encyclopedia of evaluation*. Sage, Thousand Oaks.

Mattis, J. (2018). Summary of the 2018 National Defense Strategy of the United States of America. Technical report, Department of Defense, Washington, DC. URL `https://www.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf`.

Mauw, S. and Oostdijk, M. (2005). Foundations of attack trees. In *International Conference on Information, Security and Cryptology*, Berlin Springer. doi:10.1007/11734727_17.

Miehling, E., Rasouli, M., and Teneketzis, D. (2018). A POMDP approach to the dynamic defense of large-scale cyber networks. *IEEE Transactions on Information Forensics and Security*, 13(10):2490 – 2505.

Mihaylova, L., Lefebvre, T., Bruyninckx, H., Gadeyne, K., and De Schutter, J. (2003). A comparison of decision making criteria and optimization methods for active robotic sensing. In Dimov, I., Lirkov, I., Margenov, S., and Zlatev, Z., editors, *Numerical Methods and Applications*, pages 316–324 Springer, Berlin.

Military Operations Research Society (2018). Advancing the Professionalism of Assessments. Technical report.

Miranda, E. (2008). A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628–658. doi:10.1016/j.ijar.2007.12.001.

Mohajerin Esfahani, P. and Kuhn, D. (2017). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable

reformulations. *Mathematical Programming*, 171(1-2):115–166. doi:10.1007/s10107-017-1172-1.

Monahan, G. E. (1982). State of the art – A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, 28(1): 1–16.

Montgomery, D. C. (2017). *Design and Analysis of Experiments*. John Wiley & Sons, Hoboken.

Motwani, R. and Raghavan, P. (1995). *Randomized Algorithms*. Cambridge University Press, Cambridge.

Murofushi, T. and Sugeno, M. (1991). A theory of fuzzy measures: Representations, the Choquet integral, and null sets. *Journal of Mathematical Analysis and Applications*, 159(2):532–549. doi:10.1016/0022-247X(91)90213-J.

Mushen, E. and Schroden, J. (2014). Are we winning? A brief history of military operations assessment. Technical report, CNA.

Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. (2016). *Response Surface Methodology*. John Wiley & Sons, Hoboken, 4th edition.

Nakao, H., Shen, S., and Chen, Z. (2017). Network design in scarce data environment using moment-based distributionally robust optimization. *Computers and Operations Research*, 88:44–57.

Nascimento, L. and Riella, G. (2013). Second-order ambiguous beliefs. *Economic Theory*, 52(3):1005–1037. doi:10.1007/s00199-011-0675-x.

Natarajan, K., Pachamanova, D., and Sim, M. (2009). Constructing risk measures from uncertainty sets. *Operations Research*, 57(5):1129–1141. doi:10.1287/opre.1080.0683.

Natenzon, P. (2019). Random choice and learning. *Journal of Political Economy*, 127 (1):419–457. doi:10.1086/700762.

National Performance Review (1993). Reaching public goals: Managing government for results. Technical report, Government Printing Office, Washington, DC.

Nau, R. F. (2006). Uncertainty aversion with second-order utilities and probabilities. *Management Science*, 52(1):136–145. doi:10.1287/mnsc.1050.0469.

Nesterov, Y. (2005). Excessive gap technique in nonsmooth convex minimization. *SIAM Journal on Optimization*, 16(1):235–249.

Neumaier, A. (2004). Clouds, fuzzy sets, and probability intervals. *Reliable Computing*, 10(4):249–272.

Neyshabouri, S. and Berg, B. P. (2017). Discrete optimization: Two-stage robust optimization approach to elective surgery and downstream capacity planning. *European Journal of Operational Research*, 260(1):21–40.

Nguyen, H. and Sriboonchitta, S. (2010). On Choquet integral risk measures. In Huynh, V., Nakamori, Y., Lawry, J., and Inuiguchi, M., editors, *Integrated Uncertainty Management and Applications*, pages 15–22 Springer, Berlin.

Nguyen, T. H., Jiang, A. X., and Tambe, M. (2014). Stop the compartmentalization: Unified robust algorithms for handling uncertainties in security games. *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014)*, pages 317–324.

Nguyen, T. H., Sinha, A., and Tambe, M. (2016). Addressing behavioral uncertainty in security games: An efficient robust strategic solution for defender patrols. *2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPS)*, pages 1831–1838. doi:10.1109/IPDPSW.2016.195.

Ni, Y. and Liu, Z.-Q. (2013). Bounded-parameter partially observable Markov decision processes: Framework and algorithm. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 21(6):821–863.

Nilim, A. and El Ghaoui, L. (2005). Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798.

North Atlantic Treaty Organization (2015). NATO Operations Assessment Handbook Version 3.0.

Nystrom, J. K., Robbins, M. J., Deckro, R. F., and Morris, J. F. (2018). Simulating attacker and defender strategies within a dynamic game on network topology. *Journal of Simulation*, 12(4):307–331. doi:10.1057/s41273-017-0054-0.

Office of Management and Budget (2004). Program evaluation: What constitutes strong evidence of program effectiveness? Technical report.

Office of the Chief of Naval Operations (2013). Navy Warfare Publication 5-01: Navy Planning.

Office of the Under Secretary of Defense for Policy (2017). DoD Instruction 5132.14: Assessment, Monitoring, and Evaluation Policy for the Security Cooperation Enterprise.

Oh, E. and Kim, K. E. (2011). A geometric traversal algorithm for reward-uncertain MDPs. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*.

O'Hanlon, M. E. and Campbell, J. H. (2007). Iraq Index. Technical report, The Brookings Institution, Washington, DC. URL https://www.brookings.edu/wp-content/uploads/2017/11/index20070531.pdf.

Osogami, T. (2015). Robust partially observable Markov decision process. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 106–115.

Otrok, H., Mehrandish, M., Assi, C., Debbabi, M., and Bhattacharya, P. (2008). Game theoretic models for detecting network intrusions. *Computer Communications*, 31(10):1934–1944. doi:10.1016/j.comcom.2007.12.028.

Paç, A. B. and Pnar, M. (2018). On robust portfolio and naïve diversification: Mixing ambiguous and unambiguous assets. *Annals of Operations Research*, 266(1-2):223–253. doi:10.1007/s10479-017-2619-8.

Park, J., Kim, Y. B., and Willemain, T. R. (2011). Analysis of an unobservable queue using arrival and departure times. *Computers and Industrial Engineering*, 61(3): 842–847. doi:10.1016/j.cie.2011.05.017.

Patton, M. Q. (1994). Developmental evaluation. *Evaluation Practice*, 15(3):311–319.

Patton, M. Q. (2008). *Utilization-Focused Evaluation*. Sage, Thousand Oaks, 4th edition.

Paul, C., Yeats, J., Clarke, C. P., Matthews, M., and Skrabala, L. (2015a). *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*. Rand Corporation, Santa Monica.

Paul, C., Yeats, J., Clarke, C., Matthews, M., and Skrabala, L. (2015b). *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Worked Example*. URL http://www.rand.org/content/dam/rand/pubs/research_reports/RR800/RR809z1/RAND_RR809z1.pdf.

Pazgal, A. I. and Radas, S. (2008). Comparison of customer balking and reneging behavior to queueing theory predictions: An experimental study. *Computers & Operations Research*, 35(8):2537–2548. doi:10.1016/j.cor.2006.12.027.

Perny, P., Spanjaard, O., and Storme, L. X. (2006). A decision-theoretic approach to robust optimization in multivalued graphs. *Annals of Operations Research*, 147(1): 317–341. doi:10.1007/s10479-006-0073-0.

Pineau, J., Gordon, G., and Thrun, S. (2003). Point-based value iteration: An anytime algorithm for POMDPs. *International Joint Conference on Artificial Intelligence*, pages 1025–1030.

Ponsen, M. J. V., de Jong, S., and Lanctot, M. (2011). Computing approximate Nash equilibria and robust best-responses using sampling. *Journal of Artificial Intelligence Research*, 42:575–605.

Popescu, I. (2005). A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Mathematics of Operations Research*, 30(3):632–657. doi:10.1287/moor.1040.0137.

Powell, W. B. (2007). *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. John Wiley & Sons, Hoboken.

Puterman, M. L. (2014). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, New York.

Qu, S., Meng, D., Zhou, Y., and Dai, Y. (2017). Distributionally robust games with an application to supply chain. *Journal of Intelligent & Fuzzy Systems*, 33(5): 2749–2762.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

Revels, J. (2018). BenchmarkTools. URL `https://github.com/JuliaCI/BenchmarkTools.jl`.

Romanko, O. and Mausser, H. (2016). Robust scenario-based value-at-risk optimization. *Annals of Operations Research*, 237(1-2):203–218.

Romich, A., Lan, G., and Smith, J. C. (2015). A robust sensor covering and communication problem. *Naval Research Logistics*, 62(7):582–594.

Rumsfeld, D. H. (2002). DoD News Briefing - Secretary Rumsfeld and General Myers. URL `http://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636`.

Saghafian, S. (2018). Ambiguous partially observable Markov decision processes: Structural results and applications. *Journal of Economic Theory*, 178:1–35. doi:10.1016/j.jet.2018.08.006.

Saie, C. M. and Ahner, D. K. (2018). Investigating the dynamics of nation-building through a system of differential equations. *Journal of the Operational Research Society*, 69(4):619–629. doi:10.1057/s41274-017-0256-x.

Samuelson, S. and Yang, I. (2017). Data-Driven distributionally robust control of energy storage to manage wind power fluctuations. In *2017 IEEE Conference on Control Technology and Applications*, pages 199–204.

Santos, M. C., Poss, M., and Nace, D. (2018). A perfect information lower bound for robust lot-sizing problems. *Annals of Operations Research*, 271(2):887–913. doi:10.1007/s10479-018-2908-x.

Sariddichainunta, P. and Inuiguchi, M. (2017). Global optimality test for maximin solution of bilevel linear programming with ambiguous lower-level objective function. *Annals of Operations Research*, 256(2):285–304. doi:10.1007/s10479-016-2293-2.

SAS Institute (2016). JMP. URL `https://www.jmp.com/`.

Sasaki, Y. (2017). Generalized Nash equilibrium with stable belief hierarchies in static games with unawareness. *Annals of Operations Research*, 256(2):271–284. doi:10.1007/s10479-016-2266-5.

Satia, J. K. and Lave, R. E. (1973). Markovian decision processes with uncertain transition probabilities. *Operations Research*, 21(3):728–740.

Savage, L. J. (1954). *The Foundations of Statistics*. John Wiley & Sons, New York.

Scheftelowitsch, D., Buchholz, P., Hashemi, V., and Hermanns, H. (2017). Multi-Objective approaches to Markov decision processes with uncertain transition parameters. *arXiv Preprint*.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587.

Schramm, H. C. and Gaver, D. P. (2013). Lanchester for cyber: The mixed epidemic-combat model. *Naval Research Logistics*, 60(7):599–605.

Seidl, A., Kaplan, E. H., Caulkins, J. P., Wrzaczek, S., and Feichtinger, G. (2016). Decision support: Optimal control of a terror queue. *European Journal of Operational Research*, 248(1):246–256.

Seo, K. (2009). Ambiguity and second-order belief. *Econometrica*, 77(5):1575–1605. doi:10.3982/ecta6727.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton.

Shapiro, A. (2011). A dynamic programming approach to adjustable robust optimization. *Operations Research Letters*, 39(2):83–87. doi:10.1016/j.orl.2011.01.001.

Shapiro, A. (2016). Rectangular sets of probability measures. *Operations Research*, 64(2):528–541. doi:10.1287/opre.2015.1466.

Shashua, S. D. C. and Mannor, S. (2017). Deep robust Kalman filter. *arXiv Preprint*.

Shen, D., Chen, G., Cruz, J. B., Jr., Haynes, L., Kruger, M., and Blasch, E. (2007). A Markov game theoretic data fusion approach for cyber situational awareness. In *Proceedings of SPIE 6571, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2007, 65710F*. doi:10.1117/12.720090.

Shoham, Y. and Leyton-Brown, K. (2008). *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, Cambridge.

Singh, V. V., Jouini, O., and Lisser, A. (2017). Distributionally robust chance-constrained games: Existence and characterization of Nash equilibrium. *Optimization Letters*, 11(7):1385–1405. doi:10.1007/s11590-016-1077-6.

Sinha, A., Nguyen, T. H., Kar, D., Brown, M., Tambe, M., and Jiang, A. X. (2015). From physical security to cybersecurity. *Journal of Cybersecurity*, 1(1):19–35. doi:10.1093/cybsec/tyv007.

Sinha, S. and Ghate, A. (2016). Policy iteration for robust nonstationary Markov decision processes. *Optimization Letters*, 10(8):1613–1628. doi:10.1007/s11590-016-1040-6.

Sinha, S., Kotas, J., and Ghate, A. (2016). Robust response-guided dosing. *Operations Research Letters*, 44(3):394–399.

Siniscalchi, M. (2009). Vector expected utility and attitudes toward variation. *Econometrica*, 77(3):801–855. doi:10.2139/ssrn.1030407.

Siniscalchi, M. (2011). Dynamic choice under ambiguity. *Theoretical Economics*, 6 (3):379–421. doi:10.3982/TE571.

Smets, P. (1999). Practical uses of belief functions. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, pages 612–621.

Smets, P. and Kennes, R. (1994). The transferable belief model. *Artificial Intelligence*, 66(2):191–234. doi:10.1016/0004-3702(94)90026-4.

Smith, T. and Simmons, R. (2004). Heuristic search value iteration for POMDPs. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, pages 520–527.

Smith, T. and Simmons, R. (2005). Point-Based POMDP algorithms: Improved analysis and implementation. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, pages 542–549.

Snow, A. (2010). Ambiguity and the value of information. *Journal of Risk and Uncertainty*, 40(2):133–145.

Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, 21(5):1154–1157. doi:10.1287/opre.21.5.1154.

Spaan, M. T., Veiga, T. S., and Lima, P. U. (2015). Decision-theoretic planning under uncertainty with information rewards for active cooperative perception. *Autonomous Agents and Multi-Agent Systems*, 29(6):1157–1185.

Stake, R. E. (1995). *The Art of Case Study Research*. Sage, Thousand Oaks.

Starmer, C. (2010). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, 38(2): 332–382. doi:10.1257/jel.38.2.332.

Sugeno, M. (1974). *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology.

Sun, H. and Xu, H. (2016). Convergence analysis for distributionally robust optimization and equilibrium problems. *Mathematics of Operations Research*, 41(2): 377–401. doi:10.1287/moor.2015.0732.

Szechtman, R., Kress, M., Lin, K., and Cfir, D. (2008). Models of sensor operations for border surveillance. *Naval Research Logistics*, 55(1):27–41.

Tammelin, O., Burch, N., Johanson, M., and Bowling, M. (2015). Solving heads-up limit Texas hold'em. In *International Joint Conference on Artificial Intelligence*, pages 645–652.

Tan, C. H. and Hartman, J. C. (2011). Sensitivity analysis in Markov decision processes with uncertain reward parameters. *Journal of Applied Probability*, 48(4): 954–967.

Trevizan, F. W., de Barros, L. N., and Cozman, F. G. (2007). Planning under risk and Knightian uncertainty. *International Joint Conference on Artificial Intelligence*, pages 2023–2028.

Troffaes, M. C. (2007). Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45(1):17–29. doi:10.1016/j.ijar.2006.06.001.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323. doi:10.1007/BF00122574.

Tversky, A. and Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, 101(4):547–567. doi:10.1037/0033-295X.101.4.547.

Tyler, R. W. (1942). General statement on evaluation. *The Journal of Educational Research*, 35(7):492–501. doi:10.1080/00220671.1942.10881106.

United States Air Force (2015a). Air Force Future Operating Concept. Technical report.

United States Air Force (2015b). Strategic Master Plan. Technical report.

United States Cyber Command (2018). Cyber Mission Force achieves full operational capability. URL https://dod.defense.gov/News/Article/Article/1524747/cyber-mission-force-achieves-full-operational-capability/.

United States Joint Chiefs of Staff (2015a). Operation Assessment: Multi-Service Tactics, Techniques, and Procedures. Technical report.

United States Joint Chiefs of Staff (2015b). Joint Doctrine Note 1-15: Operation Assessment.

United States Joint Chiefs of Staff (2017a). Joint Publication 5-0: Joint Planning.

United States Joint Chiefs of Staff (2017b). Joint Publication 3-0: Joint Operations.

Ure, N. K., Geramifard, A., Chowdhary, G., and How, J. P. (2012). Adaptive planning for Markov decision processes with uncertain transition models via incremental feature dependency discovery. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 99–115.

Van Parys, B. P. G., Esfahani, P. M., and Kuhn, D. (2017). From data to decisions: Distributionally robust optimization is optimal. *arXiv Preprint*.

Vardeman, S. B. (1992). What about the other intervals? *The American Statistician*, 46(3):193–197. doi:10.1080/00031305.1992.10475882.

von Neumann, J. and Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton.

Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press, Cambridge.

Wald, A. (1945). Statistical decision functions which minimize the maximum risk. *Annals of Mathematics*, 46(2):265–280.

Walker, W. E., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P., and von Krauss, M. P. (2003). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1):5–17.

Wallenius, J., Dyer, J. S., Fishburn, P. C., Steuer, R. E., Zionts, S., and Deb, K. (2008). Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Science*, 54(7):1336–1349. doi:10.1287/mnsc.1070.0838.

Walley, P. (1987). Belief function representations of statistical evidence. *Annals of Statistics*, 14(2):590–606.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, New York.

Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24(2-3):125–148.

Walraven, E. and Spaan, M. T. J. (2017). Accelerated vector pruning for optimal POMDP solvers. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 3672–3678.

Wang, C. and Guo, P. (2017). Behavioral models for first-price sealed-bid auctions with the one-shot decision theory. *European Journal of Operational Research*, 261 (3):994–1000. doi:10.1016/j.ejor.2017.03.024.

Wang, W., Sun, H. J., and Wu, J. J. (2015a). Robust user equilibrium model based on cumulative prospect theory under distribution-free travel time. *Journal of Central South University*, 22(2):761–770.

Wang, X., Chen, J., Dutta, A., and Chiang, M. (2015b). Adaptive video streaming over whitespace: SVC for 3-tiered spectrum sharing. In *2015 IEEE Conference on Computer Communications*, pages 28–36.

Wang, X., Fan, N., and Pardalos, P. M. (2018). Robust chance-constrained support vector machines with second-order moment information. *Annals of Operations Research*, 263(1-2):45–68. doi:10.1007/s10479-015-2039-6.

Wang, Z., Boularias, A., Mülling, K., and Peters, J. (2011). Balancing safety and exploitability in opponent modeling. In *25th AAAI Conference on Artificial Intelligence (AAAI-11)*, pages 1515–1520.

Weber, M. (1987). Decision making with incomplete information. *European Journal of Operational Research*, 28(1):44–57. doi:10.1016/0377-2217(87)90168-8.

Weiss, C. H. and Mark, M. M. (2006). The oral history of evaluation part IV: The professional evolution of Carol Weiss. *American Journal of Evaluation*, 27(4):475–483.

Weng, P., Qiu, Z., Costanzo, J., Yin, X., and Sinopoli, B. (2017). Optimal threshold policies for robust data center control. In *Lecture Notes in Electrical Engineering*, pages 104–114.

Whalen, T. (1984). Decisionmaking under uncertainty with various assumptions about available information. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-14(6):888–900. doi:10.1109/TSMC.1984.6313316.

White, C. C. (1991). A survey of solution techniques for the partially observed Markov decision process. *Annals of Operations Research*, 32(1):215–230.

White, C. C. and Eldeib, H. K. (1986). Parameter imprecision in finite state, finite action dynamic programs. *Operations Research*, 34(1):120–129.

White, C. C. and Eldeib, H. K. (1994). Markov decision processes with imprecise transition probabilities. *Operations Research*, 42(4):739–749.

Whitt, W. (1972). Embedded renewal processes in the GI/G/s queue. *Journal of Applied Probability*, 9(3):650–658.

Wholey, J. S., Hatry, H. P., and Newcomer, K. E., editors (2010). *Handbook of Practical Program Evaluation*. John Wiley & Sons, San Francisco, 3rd edition.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. URL `http://ggplot2.org`.

Wiesemann, W., Kuhn, D., and Rustem, B. (2013). Robust Markov decision processes. *Mathematics of Operations Research*, 38(1):152–183.

Wiesemann, W., Kuhn, D., and Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376. doi:10.1287/opre.2014.1314.

Williams, A. P. and Morris, J. C. (2009). The development of theory-driven evaluation in the military: Theory on the front line. *American Journal of Evaluation*, 30(1): 62–79. doi:10.1177/1098214008329522.

Wolff, E. M., Topcu, U., and Murray, R. M. (2012). Robust control of uncertain Markov decision processes with temporal logic specifications. In *IEEE 51st Annual Conference on Decision and Control*, pages 3372–3379.

Wong, K. Y., Lalmazloumian, M., Wong, K. Y., Govindan, K., and Kannan, D. (2016). A robust optimization model for agile and build-to-order supply chain planning under uncertainties. *Annals of Operations Research*, 240(2):435–470.

Woodford, M. (2014). Stochastic choice: An optimizing neuroeconomic model. *American Economic Review*, 104(5):495–500.

Wozabal, D. (2012). A framework for optimization under ambiguity. *Annals of Operations Research*, 193(1):21–47. doi:10.1007/s10479-010-0812-0.

Wray, K. H., Kumar, A., and Zilberstein, S. (2018). Integrated cooperation and competition in multi-agent decision-making. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*.

Xiao, H., Yang, K., Wang, X., and Shao, H. (2012). A robust MDP approach to secure power control in cognitive radio networks. In *2012 IEEE International Conference on Communications*, pages 4642–4647.

Xin, L. and Goldberg, D. A. (2015). Distributionally robust inventory control when demand is a martingale. *arXiv Preprint*.

Xu, H. and Mannor, S. (2012). Distributionally robust Markov decision processes. *Mathematics of Operations Research*, 37(2):288–300. doi:10.1287/moor.1120.0540.

Yager, R. R. (1979). Possibilistic decisionmaking. *IEEE Transactions on Systems Man and Cybernetics*, 9(7):388–392.

Yang, I. (2017a). Distributionally robust stochastic control with conic confidence sets. In *56th IEEE Conference on Decision and Control*, pages 4291–4296.

Yang, I. (2017b). A convex optimization approach to distributionally robust Markov decision processes with Wasserstein distance. *IEEE Control Systems Letters*, 1(1): 164–169.

Yang, L., Li, Y., Chen, K., and Zhou, Z. (2014). Distributionally robust return-risk optimization models and their applications. *Journal of Applied Mathematics*.

Ye, N., Somani, A., Hsu, D., and Lee, W. S. (2017). DESPOT: Online POMDP planning with regularization. *Journal of Artificial Intelligence Research*, 58:231–266.

Yin, R. K. (2009). *Case Study Research: Design and Methods*. Sage, London and Singapore.

Yost, K. A. and Washburn, A. R. (2000). LP/POMDP marriage: Optimization with imperfect information. *Naval Research Logistics*, 47(8):607–619.

Yu, P. and Xu, H. (2016). Distributionally robust counterpart in Markov decision processes. *IEEE Transactions on Automatic Control*, 61(9):2538–2543. doi:10.1109/TAC.2015.2495174.

Yu, S., Liu, Z., and Wu, J. (2016). Equilibrium strategies of the unobservable M/M/1 queue with balking and delayed repairs. *Applied Mathematics and Computation*, 290:56–65. doi:10.1016/j.amc.2016.05.049.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353. doi:10.1016/S0019-9958(65)90241-X.

Zadeh, L. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1):3–28. doi:10.1016/0165-0114(78)90029-5.

Zadeh, L. (1984). Review of a mathematical theory of evidence. *AI Magazine*, 5(3): 81. doi:10.1609/aimag.v5i3.452.

Zadeh, L., Bellman, R., and Robbins, H. (2005). Toward a generalized theory of uncertainty (GTU) – An outline. *Information Sciences*, 172(1-2):1–40. doi:10.1016/j.ins.2005.01.017.

Zhang, N. L. and Liu, W. (1996). Planning in stochastic domains: Problem characteristics and approximation. Technical Report HKUST-CS96-31, Hong Kong University of Science and Technology.

Zimmermann, H. J. (1978). Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets and Systems*, 1(1):45–55. doi:10.1016/0165-0114(78)90031-3.

Zinkevich, M., Johanson, M., Bowling, M., and Piccione, C. (2008). Regret minimization in games with incomplete information. *Advances in Neural Information Processing Systems*, pages 1729–1736.

Zonouz, S. A., Khurana, H., Sanders, W. H., and Yardley, T. M. (2014). RRE: A game-theoretic intrusion response and recovery engine. *IEEE Transactions on Parallel and Distributed Systems*, 25(2):395–406. doi:10.1109/CRIS.2009.5071485.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
12–09–2019	Doctoral Dissertation	September 2016 — September 2019

4. TITLE AND SUBTITLE

Operational Decision Making Under Uncertainty: Inferential, Sequential, and Adversarial Approaches

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Keith, Andrew J., Capt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-DS-19-S-041

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Jill Morrissett, jill.s.morrissett.civ@mail.mil
United States Strategic Command
4048 Higley Rd
Dahlgren, VA 22448

10. SPONSOR/MONITOR'S ACRONYM(S)

STRATCOM

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Modern security threats are characterized by a stochastic, dynamic, partially observable, and ambiguous operational environment. This dissertation addresses such complex security threats using operations research techniques for decision making under uncertainty in operations planning, analysis, and assessment. First, this research develops a new method for robust queue inference with partially observable, stochastic arrival and departure times, motivated by cybersecurity and terrorism applications. In the dynamic setting, this work develops a new variant of Markov decision processes and an algorithm for robust information collection in dynamic, partially observable and ambiguous environments, with an application to a cybersecurity detection problem. In the adversarial setting, this work presents a new application of counterfactual regret minimization and robust optimization to a multi-domain cyber and air defense problem in a partially observable environment.

15. SUBJECT TERMS

Decision making under uncertainty, operation assessment, operation planning, ambiguity, queueing, Markov decision process, robust optimization, counterfactual regret minimization, game theory, machine learning

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Darryl K. Ahner, AFIT/ENS
U	U	U	UU	292	19b. TELEPHONE NUMBER *(include area code)* (937) 255-6565, x4708; darryl.ahner@afit.edu